Materials Advances

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: A. Jalilov and A. N. Duman, *Mater. Adv.*, 2024, DOI: 10.1039/D4MA00505H.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/materials-advances

View Article Online
DOI: 10.1039/D4MA00505H

Machine Learning for Carbon Dots Synthesis and Applications

Almaz S. Jalilov† and Ali Nabi Duman*,‡

†*Department of Chemistry, King Fahd University of Petroleum and Minerals, Dhahran*

‡*Department of Mathematics and Statistics, University of Houston-Downtown, Houston*

E-mail: dumana@uhd.edu

Abstract

One of the hottest topics in the nanoparticles research right now is carbon dots(CDs). In order to be used in applications like medical imaging and diagnostics, pharmaceutics, optoelectronics, and photocatalysis, CDs must be synthesized with carefully controlled properties. This is often a tedious task due to the fact that nanoparticle syntheses frequently involve multiple chemicals and are carried out under complex experimental conditions. The emerging data-driven methods from artificial intelligence (AI) and machine learning (ML) provide promising tools to go beyond time-consuming, and laborious trial-and-error approach. In this review, we focus on the recent uses of ML accelerating exploration of CD chemical space. Future applications of these methods address current limitations in CD synthesis expanding the potential uses of these intriguing nanoparticles.

Introduction

Due to their distinctive advantages, such as simple synthesis, long-term photo and colloidal stability, biocompatibility, biodegradability, non/low toxicity, low cost, tunable photolu-

View Article Online
DOI: 10.1039/D4MA00505H

minescence, and good dispersibility, carbon-based nanomaterials, particularly carbon dots (CDs) have been one of the most studied materials in recent years.[1–8] The favorable characteristics make CDs susceptible of applications in biosensing and bioimaging,[9–11] cancer research,[12] drug delivery,[13] visible light communication,[14,15] and optoelectronic devices.[16–19] A base carbon core with chemical functional groups attached or modified on the surface makes up the core-shell-like structure of CDs. The surface generally consists of some common functional groups, such as amino, epoxy, carbonyl, aldehyde, hydroxyl, and carboxylic acid, while the carbon core structure consists of sp^2 and sp^3 carbon atoms.[20–22] The additional molecular structures which are essential to their features make CDs extremely complex.

CDs can be synthesized by utilizing bottom-up or top-down methods.[16,23] In the top-down methods, large carbon materials are cut into small carbon structures smaller than 10 nm. The demanding physical procedures to break down the carbon materials (e.g., graphite, graphene oxide, carbon nanotubes, activated carbon, soot) involve laser ablation, arc discharge and nanolithography under unfavorable conditions such as strong oxidants, concentrated acids, and high temperatures.[24–33] The more adaptable and accessible bottom-up methods usually includes ultrasound synthesis, chemical oxidation, room temparature method, hydrothermal and solvothermal processing of relatively small molecular precursors.[34] Although these methods may include high temperatures/pressures, long reaction times, or toxic solvents, use of microwaves in solvothermal synthesis partially solves these issues by reducing the reaction time and the amount of solvents.[35–40] The room temperature method is another advantageous technique because it does not require complicated machinery or harsh synthesis conditions, making it environmentally friendly and sustainable.[41–43] Hence, its simple setup, low-cost and accessibility to wide variety of percursors makes bottom-up methods more favorable over the top-down methods.

CDs can be divided into four main classes acccording to their carbon core structure, surface functionalities, and performance features: (i) graphene quantum dots (GQDs), (ii) carbon nanodots (CNDs), (iii) carbon quantum dots (CQDs), and (iv) carbonized polymer

View Article Online
DOI: 10.1039/D4MA00505H

dots (CPDs).[44,45] GQDs are usually synthesized by using top-down approach, while preparation of CNDs and CQDs are mainly done by using bottom-up methods.[46–48] A variety of models (e.g polycyclic aromatic hydrocarbons, molecular fluorophores, or sp2/sp3 hybrid spherical structures) are used to explain different structures of CDs.[49–52]

Due to the necessary high temparatures during bottom-up sythesis of CDs, multiple reaction pathways occur while forming considerable amount of by-products. Along with the irregular mass transfer, low reproducibility is also common as reported in earlier studies.[35,53] One solution to optimize the target properties is to scan large experimental synthesis conditions including reaction temperature, the mass of precursor, ramp rate, and reaction time. However, high complexity of the extracted data, repetitive experimental procedures, and the lack of predictibility make this scan very time consuming to achieve ideal results. .For example, it is still unclear how the CQDs emit their fluorescence because it is a very complicated process. It is customary to analyze the pH-dependent photoluminescence (PL) spectra of CQDs at a fixed excitation and ignore all other potential excitations; however, this method only allows for the extraction of a portion of the available data.[39,54–60] On the other hand, the complexity of data analysis methods can rise along with the number of PL measurements. Similarly, current CDs reported in the literature were frequently prepared optimally by controlling one reaction parameter and fixing other reaction factors, while not considering complex relationship between reaction parameters during CD synthesis.Therefore, there is a need to employ methods which accelerate the screening of the necessary parameters in order to create CDs with enhanced features and applications.

Quantum mechanics methods such as density functional theory (DFT) is a reliable computational solution to search reasonably designed parameter space.[61,62] These semi-empirical approaches can be used to explore electronic structure and chemical reactivity of CDs.[63–67] Density-functional-based tight binding (DFTB) is another semi parametric method which approximates DFT in a tight binding framework.[68–71] DFTB requires less empiriral parameters and computationally more efficient than DFT. Mechanism of graphene formation and

View Article Online
DOI: 10.1039/D4MA00505H

single-walled carbon nanotube nucleation are examples studied using DFTB.[72,73] However, these semi-empirical methods are computationally too costly for a large search space. The alternative approaches to reduce the entire search space include optimization and gradient based algorithms. The accuracy and computational performance of these methods depend on the initially determined parameters; hence, they might return different results from the different initial values and potentially end up in local minima.

Data-driven approaches based on machine learning (ML) algorithms provide an alternative to the above mentioned computational methods for the description of structure and properties of CDs. As a branch of artificial intelligence, ML employs statistical and probabilistic methods to learn from a given dataset by optimizing a performance measures for particular tasks.[74,75] Certain methods have the ability to detect the relationship (correlations/inference) between input variables and target variable. Instead of screening entire parameter space, ML methods learn the hidden patterns using limited amount of data. These trained algorithms are later generalized to predict the target variables from previously unseen input variables. As a result of increasing amount of experimental data, and accessible computational power; ML have successful applications in variety of fields including image/speech recognition, cancer research, chemical synthesis, protein structure prediction.[76–84]

In material science, ML has attracted a lot of interest in applications such as material discovery, material structure/property prediction, performance optimization, and acceleration of the protocols for nanoparticle synthesis.[85–95] Using ML, the reaction parameters and their effects on the nanoparticle synthesis can be revealed objectively,[96,97] and the synthesis process can be made more efficient by choosing appropriate evaluation criteria including shape, size, polydispersity, and surface chemistry.[98] ML accelerates not only the experimental protocols but also search of new semiconductor, metal, carbon-based, and polymeric nanoparticles with superior features requiring low computational cost.[8] The large amount of data needed for ML algorithms can be obtained using computational or experimental methods. Numerous databases such as the Materials Project, Automatic Flow for Materials

View Article Online
DOI: 10.1039/D4MA00505H

Discovery, Open Quantum Materials Database, Novel Materials Discovery make it possible to access a lot of materials' data in addition to using computer simulations to generate it.

In particular, the use of ML in the field of CD has generated a lot of interest in research in recent years. The proper adjustment of a variety of variables, including precursors, temperature, and reaction time, is necessary for the successful preparation of CDs. It is simple to use these elements as input parameters in ML, which is trained with the available experimental data and generates accurate new predictions. Therefore, the addition of ML can aid in the relationship between precursors and desired properties, which may result in the formation of a design principle for further study and significantly shorten the synthesis cycle and lower the cost of CDs.

Many outstanding reviews on the applications of computational and ML methods to the nanoparticle synthesis have been written.[98–106] Although some of them concentrates CD synthesis, they partially cover the developments of ML methods along with the experimental techniques.[99] The more general reviews include CDs as a subcategory of nanoparticles,[98,100] quantum dots,[103,104] graphene-based[101,105] or polymer-based[106] materials. Theoretical methods such as quantum mechanics and/or molecular mechanics approaches applied to CDs is also available in the literature.[102] To the best of our knowledge, a thorough review of ML applications specifically for CD synthesis is lacking. In this review, we outline the primary ML algorithms in the context of CDs research, discuss recent studies on ML applications for CD synthesis, and enumerate potential future directions for this rapidly expanding field of study. The related papers are listed in Table 1.

Linear Regression

Numerous machine learning algorithms have been devised for diverse learning scenarios, encompassing unsupervised, semi-supervised, and supervised learning. In scientific and engineering contexts, supervised learning, also known as predictive modeling, is widely favored.

Table 1: Recent studies on applying ML algorithms to carbon dots research.

ML models	Input	Output	Samples	Refs.
MLR, Poly. Reg.	Microstructural features	Thermoelectric performance	322	[107]
Lin. reg., Poly. reg.	Synthesis process parameters	UV-visible and PL spectra	44	[108]
MLR, KNN	Synthesis process parameters	PLQY, PL peak position	227	[109]
MLR	PL characteristics	Temperature sensing accuracy	121	[110]
Random Forest	Reaction parameters	Emission wavelength, stokes shift, PLQY	480	[111]
Log. reg., KNN, SVM	CDs fluorescence sensor array	Protein classification	48	[112]
CNN	Synthesis process parameters	spectral properties and FL colors	170	[113]
ANN	Synthesis process parameters	Color classification, emmission wavelenth	407	[114]
ANN	CDs fluorescence variation maps	Aminoacid classification	90	[115]
Multilayer Perceptron	Synthesis process parameters	PLQY	30	[116]
CNN	Emission/PL decay data of CDs	Ethanol content prediction	597	[117]
XGBoost	Synthesis process parameters	PLQY	391	[82]
XGBoost	Synthesis process parameters	PLQY	467	[118]
PCA, XGBoost	Synthesis process parameters	FL intensity, emission centers	400	[119]
XGBoost	Reaction parameters of CD catalysts	Failure/success of oxidation of C–H bonds	652	[120]
GBDT	Biochar preparation parameters	Fluorescence quantum yield	480	[121]
Random Forest	Precursor combinations	PL wavelength intensity	202	[122]
Random Forest, GA	Synthesis process parameters	Corrosion inhibition efficiency	102	[123]
PCA, MCR, NMF	Wavelengths, pH	Unsupervised clustering	401	[124]
LDA, SVM	CDs fluorescence sensor array	Tetracycline classification	92	[125]

Materials Advances

View Article Online
DOI: 10.1039/D4MA00505H

View Article Online
DOI: 10.1039/D4MA00505H

Of all the supervised learning methods, linear regression is the most fundamental, having been extensively studied and applied due to its simplicity and high interpretability. Given input variables $\mathbf{x} = (x_1, \ldots, x_p)$, the output variable $\mathbf{y}$ is predicted by

$$\mathbf{y} = \mathbf{x}^T \beta$$

. The most common objective function to determine the coefficients β is residual sum of squares:

$$RSS(\beta) = \sum_{1}^{N} (y_i - x_i^T \beta),$$

where N is the number of data points. Basic expansions of input variables, such as $x_2 = x_1^2$, $x_3 = x_1^3$ leads to polynomial regression. One can also further modify the linear regression models by shrinking the estimated coefficients to zero. Using methods like ridge regression and Least Absolute Shrinkage and Selection Operator (LASSO), shrinkage the coefficients reduces the weight of irrelevant input variables to resulting in more interpretable and accurate models.

Applying multiple linear regression(MLR), Armida et al. explored the relationship between the size, dimensionality, concentration, doping and other microstructural features of carbon dots and their thermoelectric performance.[107] The conversion efficiency of a thermoelectric material is quantified by the thermoelectric figure of merit, $ZT = S^2\sigma T/\kappa$, where S is the Seebeck coefficient, σ is the electrical conductivity, κ is the thermal conductivity, and T is the absolute temperature. MLR is performed for each ZT, σ, T and κ using 10 input variables characterizing size, dimensionality, concentration, doping and other features. The results revealed a strong negative relationship between functionalization and S, as well as a strong positive relationship between the type of carbon nanostructures and σ. Polynomial regression highlighted significant impacts of six input parameters on the Seebeck coefficient, electric conductivity σ and thermal conductivity κ, while no combination of parameters significantly affected thermoelectricity ZT.

View Article Online
DOI: 10.1039/D4MA00505H

Zhang et al. utilized linear and polynomial regression models to investigate core synthesis process parameters of B,N-GQDs (synthesis temperature, H_2O_2 additional volume, and synthesis time).[108] The models are trained using optical properties of B,N-GQDs derived from UV-visible and PL spectra (i.e. 675/500 peak intensity ratio and PLQY). While the authors performed other complex models such as bagging regression, random forest regression, lasso regression, and ridge regression, the highest R^2-score is obtained by polynomial model of degree 7 ($R^2 = 0.9860$). Polynomial and linear regression models pointed out that high H_2O_2 additional volume, low synthesis temperature, and appropriate synthesis time in the selected process conditions contribute to achieving a high 675/500 peak intensity ratio (see Figure 1).

Tuchin et al. analysed a dataset on the synthesis parameters and optical characteristics of carbon dots focusing on their optical behavior within the red and near-infrared wavelengths.[109] A predictive model using multiple linear regression has been developed to forecast the spectral attributes of these carbon dots. This model's validity was confirmed by comparing its predictions with the actual optical properties observed in carbon dots synthesized across three distinct laboratories.

Doring et al. applied a multiple linear regression model that combines steady-state and time-resolved luminescence data from carbon dots to enhance temperature sensing accuracy to 0.54 K.[110] This research illustrates the significant advancements in temperature sensing using optical probes through multidimensional machine learning techniques.

Several machine learning algorithms are dedicated to classification tasks. Logistic regression considered under the umbrella of generalized linear model is specifically designed for predicting probabilities associated with discrete (categorical) variables. The probability $p(\mathbf{x})$ of a sample belonging to a particular category is expressed as:

$$p(\mathbf{x}) = \frac{e^{\mathbf{x}^T\beta}}{1 + e^{\mathbf{x}^T\beta}}.$$

View Article Online
DOI: 10.1039/D4MA00505H

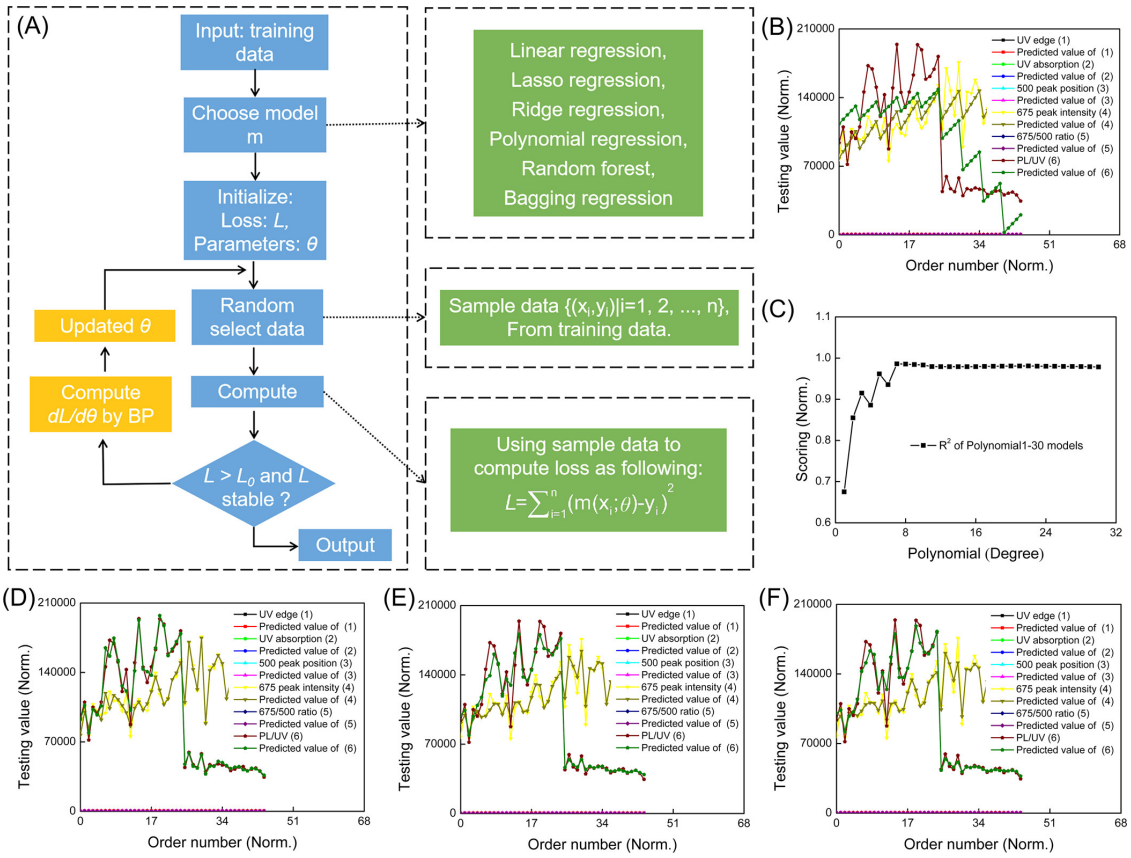

Figure 1: Machine learning-assisted evaluation of B,N-GQDs. (A) Schematic of machine learning-assisted evaluation of optical properties of B,N-GQDs. Optical properties of B,N-GQDs in varied synthesis conditions and corresponding predicted value sets with (B) linear regression, (C) polynomials 1-30, (D) polynomial regression 7, (E) bagging regression, and (F) random forest regression. The R2 scores of linear regression, polynomial regression 7, bagging regression, and random forest regression models are 0.6751, 0.9860, 0.9473, and 0.9469, respectively. Reprinted with permission from.[108] Copyright 2022 American Chemical Society.

Here, $\mathbf{x}$ represents the input variables and β denotes the coefficient vector, which is determined through the optimization of an objective function. The training process for logistic regression involves minimizing objective functions such as LBFGS (Limited-memory Broyden-Fletcher-Goldfarb-Shanno), Newton, and Stochastic Gradient Descent. Pandit et al. presented a biomolecular sensor utilizing a CD array for the detection of proteins in both buffer and human serum.[112] They anticipated that introducing analytes to CDs featuring diverse surface functionalities would induce a distinctive fluorescence change pattern. This pattern could subsequently be examined using machine learning techniques including logistic

View Article Online
DOI: 10.1039/D4MA00505H

regression. They trained their models with response of CD arrays of 48 examples from 8 classes of protein. Logistic regression, in conjunction with three distinct machine learning algorithms (namely kNN, Gradient-Boosted trees, and Support Vector Machine), attained a perfect accuracy of 100% on the test set comprising 24 unidentified samples. Hence, the effectiveness of employing machine learning algorithms for the pattern recognition of fluorescence signals from the array has been successfully demonstrated.

Artificial Neural Networks

Linear models excel when a linear connection exists between input and output variables; however, their accuracy diminishes in the presence of nonlinear interactions between variables. Artificial neural networks (ANNs) are employed to surmount these constraints. ANNs are characterized by multiple hidden layers, excluding input and output layers. Each hidden layer consists of numerous neurons, employing linear regression with a nonlinear activation function. Notably, the hidden layers feature complete connectivity between neurons of adjacent layers. It is proven that any continous function can be approximated by ANNs with one hidden layer. Convolutional neural networks (CNNs) stand as a prevalent architecture within artificial neural networks (ANNs), finding particular application in the analysis of images and videos. At the heart of the CNN lies the convolution layer, featuring multiple convolution filters. Each filter undergoes convolution with the input from the preceding layer, producing feature maps subsequently utilized as input for the following layer. The streamlined interconnection between layers contributes to the computational efficiency of CNNs, enabling them to outperform basic ANN models, particularly in tasks such as image classification.

ANNs have found application in various carbon dots studies. Wang et al. have used a convolutional neural network (CNN) model tailored for the predicting carbon dots' optical characteristics such as spectral properties and fluorescence (FL) colors under ultraviolet (UV)

View Article Online
DOI: 10.1039/D4MA00505H

irradiation.[113] The model is trained with CD synthesis features (precursor, mass, temperature, solvent, and reaction time) from 170 prototypical studies. The output layer is a feature vector that indicates spectral properties and FL color under UV irradiation. Subsequently, CDs with distinct emission properties were synthesized, and their experimental data was compared with the predicted outcomes from the trained model. These synthesized CDs were employed in cell imaging, demonstrating good performance. These findings suggest that the implementation of CNN can assist researchers in achieving effective CD design without the need for extensive manual processes.Within the same study, alternative classification models, such as support vector machines, k-nearest neighbors, random forests, decision trees, and extreme gradient boosting, demonstrated inferior performance compared to CNN. These outcomes underscore the significant potential of CNNs in guiding the synthesis of CDs.

Senanayake et al. conducted a parallel study, employing ANN, to characterize the influence of synthesis parameters on and make predictions for the emission color and wavelength of CDs.[114] The machine analysis indicated that the selection of the reaction method, purification method, and solvent is more closely correlated with CD emission characteristics compared to factors like reaction temperature or time, which are often adjusted in experimental settings. A total of 407 data examples were gathered from the literature, with 379 of them constituting the training database. The remaining 28 data examples were reserved as an external test set to validate the model. The color prediction from the classification model, which does not include reaction temperature and time as features, attained a training accuracy value of 0.94. The accuracy of emission prediction is enhanced from MAE = 38.4 to 25.8 when a combination of both classification and regression methods is employed. To overcome the limitations associated with a small dataset in a ANN model, the authors used ANN k-ensemble model which outpeformed XGBoost,K-nearest neighbor (KNN), and support vector machine (SVM). The hybrid models employed a two-step approach: initially, a classification model was utilized to predict the color, and subsequently, this predicted color (combined with the actual color during training) served as an input to predict the emission

View Article Online
DOI: 10.1039/D4MA00505H

wavelength using a regression machine learning model.The tools developed in this study, particularly the hybrid models, are expected to be valuable in predicting the emission of novel carbon dots (CDs). This approach allows for the selection of promising reaction examples from the model, streamlining the synthesis of CDs with specific colors and significantly reducing the effort required in the optimization process.

In another classification problem, Tuccito et al. employed CD fluorescence as a nanochemosensors to detect different aminoacids.[115] The modification of CD surfaces can alter fluorescence properties, including emission intensity, excitation, and emission wavelengths. In this study, carboxyl groups on nanoparticle surfaces were activated and subsequently reacted with various amino acids. The nanochemosensors demonstrated the ability to distinguish among amino acids within a mixture, showcasing their potential in complex amino acid analyses. ANN was trained with fluorescence variation maps of activated CDs to predict if the aminoacid alanine (ALA) or not alanine. The resulting model had 0.8 sensitivity and 0.91 specificity. These discoveries will contribute to the advancement of cost-effective nanochemosensors for investigating specific diseases that are presently diagnosed through basic amino acid detection methods.

In a regression task, Pudza et al. applied multilayer perceptron (MLP) to predict photoluminescent quantum yield (PLQY) of fluorescent carbon dots synthesized from tapioca powder.[116] The training data ($n = 30$) was collected from the experiments. MLP trained with temperature, time, dosage and solvent ratio predicted PLQY with high accuracy. The optimization and prediction processes have yielded sustainable, efficient, and reliable fluorescent carbon dots. This approach not only saves energy within a manageable timeframe but also reduces the required dosage while maintaining an optimal quality output.

Doring et al. applied CNN and deep neural networks on emission/PL decay data of CDs to improve ethanol content determination in ethanol/water mixtures ($n = 578$) as well as in alcohol-containing beverages ($n = 19$).[117] The models are trained by PL excitation/emission maps, PL decay spectra, and extracted features (i.e. PL intensities, PL peak positions,

View Article Online
DOI: 10.1039/D4MA00505H

and PL lifetimes) to predict the ethanol content. The utilization of time-resolved spectral information (PL decays and lifetimes) as input for CNN enables more accurate prediction of ethanol content compared to steady-state emission data. Using entire optical spectra, namely PL decays and PL excitation/emission maps, advanced deep learning models were demonstrated applicability in the analysis of beverages. In contrast to CNN models with only a few predictor variables, which struggled due to autofluorescence of the beverages, advanced deep learning models enabled better predictions of ethanol contents. Although CDs serve as excellent candidates for showcasing deep learning in optical sensing, the methods outlined in this study hold promise for enhancing chemical sensing across a range of luminescent materials (see Figure 2).

Gradient Boosting

Traditional gradient boosting techniques have long served as stalwarts in the realm of machine learning, providing a robust framework for constructing powerful predictive models. The foundational concept behind gradient boosting involves the sequential training of weak learners, such as decision trees, with each subsequent learner aiming to correct errors made by the ensemble of preceding ones. This iterative process enables the algorithm to progressively refine its predictive accuracy, making gradient boosting a popular choice for regression and classification tasks. Despite its success, traditional gradient boosting is not without its limitations. The absence of explicit regularization mechanisms can lead to overfitting, especially in the presence of noisy or high-dimensional datasets. Recognizing these challenges, the advent of Extreme Gradient Boosting (XGBoost) marked a significant evolution in the field, addressing these limitations and introducing innovations that have propelled it to the forefront of machine learning algorithms.[126]

Extreme Gradient Boosting (XGBoost), a powerful ensemble learning algorithm, has emerged as a dominant force in the realm of machine learning, demonstrating remarkable

View Article Online
DOI: 10.1039/D4MA00505H

success across various domains. Developed as an extension of traditional gradient boosting techniques, XGBoost has garnered widespread popularity due to its efficiency, scalability, and superior predictive performance. At its core, XGBoost operates by sequentially training a series of weak learners, typically decision trees, and iteratively refining their predictive capabilities. Unlike traditional gradient boosting, XGBoost incorporates a regularization term and employs a second-order Taylor expansion to optimize the objective function, enhancing its ability to capture complex patterns within the data.

One of the defining features of XGBoost is its versatility, making it applicable to both regression and classification tasks. The algorithm excels in handling large datasets and high-dimensional feature spaces, showcasing robustness in the face of noisy or missing data. Moreover, XGBoost provides a comprehensive set of hyperparameters that can be fine-tuned to accommodate diverse modeling scenarios, fostering adaptability to different applications. The algorithm's success is further underscored by its ability to balance bias and variance, mitigating overfitting and ensuring generalizability across unseen data. As a result, XGBoost has become a method of choice in various fields, ranging from finance and healthcare to image processing and natural language processing, showcasing its broad utility and effectiveness in extracting meaningful patterns from complex datasets.

XGBoost has demonstrated considerable efficacy in numerous CD studies. Han et al. reported a machine learning-assisted approach for synthesizing highly fluorescent CDs using the hydrothermal route.[82] XGBoost outperformed multilayer perceptron, support vector machine, and Gaussian process regressor in predicting the QY using five input variables: volume of ethylenediamine, mass of precursor, reaction temperature, ramp rate and reaction time. The data was collected from 391 experiments with different combinations of growth parameters, and respective QYs ranging from 0 to 1. XGBoost unveiled a noteworthy correlation between outstanding optical properties and the mass of the precursor and the volume of the alkaline catalyst. This observation aligns well with experimental findings. The methodology introduced in this study serves as a foundational step toward the advancement of artificial

View Article Online
DOI: 10.1039/D4MA00505H

intelligence techniques for the analysis and optimization of material preparation methods (see Figure 3).

Tang et al. developed a regression model to improve the PLQY of carbon quantum dots (CQDs) grown through hydrothermal methods.[118] Six hydrothermal parameters were identified as input features: pH value (pH), reaction temperature (T), reaction time (t), mass of precursor A (M), ramp rate (Rr), and solution volume (V). A total of 467 experimental records were used with different growth parameters and respective PLQY ranging from 0 to 1. In order to best infer PLQY from the features, several regression algorithms are evaluated with nested cross validation, including XGBoost regressor, support vector machine regressor, Gaussian process regressor. XGBoost demonstrates superior performance, surpassing the other algorithms by a significant margin, as indicated by its R^2 value of 0.8402. The most critical factor influencing the PLQY is shown to be the pH value, with reaction temperature and reaction time following closely in significance. The trained XGBoost model is then employed to predict the PLQY for a vast array of 1,555,840 potential synthesis conditions generated from various combinations. Eleven synthesis conditions are recommended by the model attributed to their highest predicted PLQY. Subsequent experiments conducted in the laboratory yielded a remarkably high photoluminescence quantum yield (PLQY) of 55.5%. This achievement is particularly noteworthy given the ultra-low heteroatom doping precursor ratio employed, making it one of the highest reported PLQY values under such conditions. The findings support the promising potential of ML in optimizing and expediting the material synthesis process. This endorsement suggests that ML has the capability to facilitate the development of advanced inorganic materials, contributing to practical applications through reduced processing time and enhanced material properties.

Hong et al. utilized the XGBoost model for predicting the maximum fluorescence (FL) intensity and emission centers of CDs synthesized under room temperature conditions using p-benzoquinone (PBQ) and ethylenediamine (EDA) as starting materials.[119] They successfully synthesized a variety of CDs with tailored optical properties. These CDs were

View Article Online
DOI: 10.1039/D4MA00505H

effectively employed for applications such as detecting Fe^{3+}, facilitating sustained drug release, enabling whole-cell imaging, and contributing to the preparation of poly(vinyl alcohol) (PVA) films. The input dataset comprises four hundred types of CDs prepared under different reaction conditions, encompassing the mass of p-benzoquinone(V_{EDA}), volume of ethylenediamine(V_{EDA}), reaction duration, and solvent types. For output, the predicted target variables are the FL intensity and the location of emission centers. Principal Component Analysis (PCA) was employed to create new variables characterized by relative independence. Subsequently, PC1 and PC2 were utilized as novel input features for the training of the model. XGBoost showed superior performance compared to K-nearest neighbor, decision trees, random forest, support vector machine and convolutional neural network. Leveraging the significant features and parameters (i.e.V_{EDA} and M_{PBQ}) extracted from the XGBoost model, the authors successfully fabricated a series of novel carbon dots (CDs) with customizable fluorescence (FL) intensity and emission center properties. This study demonstrates that the XGBoost algorithm, as a machine learning approach, is effective in identifying crucial factors in CD synthesis. It provides chemists with a rapid and reliable means to access optimal reaction parameters for synthesizing desired CDs (see Figure 4).

Using ML, Wang et al. successfully predicted and synthesized metal-free CD homogeneous catalysts for the oxidation of C–H bonds.[120] The dataset for cyclohexane oxidation was compiled from literature sources and laboratory notebooks, comprising a total of 652 entries. This dataset consists of 113 positive samples (17.3%) and 539 negative samples (82.7%). The boundary between success and failure in this context is characterized by achieving a 10% conversion of cyclohexane and a 70% selectivity towards the production of apidic acid(AA). The input features are selected as O (content of oxygen), Mw (weight-average molecular weight of the nonmetal catalyst), G (O_2 or not), p (homogeneous catalysis or heterogeneous catalysis), T (catalytic temperature), P (pressure), and t (reaction time). Out of the four classical models considered (Multilayer Perceptron, Naive Bayes, SVM, and XGBoost), the XGBoost model was chosen due to its high performance. The analysis of feature importance

View Article Online
DOI: 10.1039/D4MA00505H

derived from the XGBoost indicates that molecular weight (Mw) takes precedence over other features. The order of importance follows Mw, followed by O, T, P, and t. Subsequently, the established XGBoost model is employed to apply the unexplored conditions, predicting the probability of success or failure. All predictions align with actual outcomes of "success" in the conducted true experiments, affirming the accuracy of the model. This study distinctly illustrates a novel approach to C−H bond activation, employing metal-free CDs as quasi-homogeneous catalysts.

Chen et al. explored the relationship between biochar preparation parameters and fluorescence quantum yield of CDs in biochar, employing six machine learning models including decision trees (DT), random-forest (RF), gradient- boosting decision-trees (GBDT), extra-trees(ET), K- nearest-neighbor regression (KNN), and XGBoost(XGBoost), where the dataset consisted of 480 samples.[121] The input parameters for the biochar production experiment were determined, encompassing the type of farm waste, as well as characteristics such as cellulose, hemicellulose, lignin, ash, moisture, nitrogen (N), carbon (C), and carbon-to-nitrogen ratio (C/N) contents of the samples. Additionally, parameters related to the pyrolysis process, including pyrolysis temperature (T) and residence time (t), were considered.The GBDT model had the best performance among the other models; as GBDT exhibit resilience to missing values and outliers, are less susceptible to the impact of extreme values, and demonstrate effectiveness in handling high-dimensional sparse data. It was identified that four features pyrolysis temperature, residence time, nitrogen (N) content, and carbon-to-nitrogen (C/N) ratio, had the most significant impact on enhancing the accuracy of QY predictions. The methodology introduced in this study can serve as a foundation for the advancement of new techniques leveraging artificial intelligence for the analysis and prediction of CDs generated in the process of biochar production.

View Article Online
DOI: 10.1039/D4MA00505H

Random Forest

Random Forest (RF), a versatile and robust machine learning algorithm, has emerged as a cornerstone in data-driven decision-making across diverse scientific and industrial domains. Born out of the ensemble learning paradigm, this algorithm is particularly well-suited for applications where the accuracy and reliability of predictions are paramount. The metaphorical "forest" comprises a multitude of decision trees, each contributing its unique insights to the collective wisdom of the algorithm. As a result, Random Forest is known for its resilience against overfitting and its ability to produce accurate and stable predictions. The integration of Random Forest algorithms in the study of carbon dots opens up new avenues for predicting and understanding their behavior.

Chen et al. explored the relationship between reaction parameters and the photoluminescence characteristics of CDs, achieving controllable synthesis of multi-color CDs with the aid of ML.[111] Five input parameters are used, including varied precursor types and quantities such as p-phenylenediamine with urea, p-phenylenediamine with citric acid, diverse solvent types (anhydrous ethanol, water, N, N-dimethylformamide), along with reaction time and temperature. 270 experiments with different parameter combinations is conducted to feed the ML algorithms. The 3D fluorescence spectra (maximum emission wavelength, stokes shift) and fluorescence quantum yield were used as output variables. The RF model demonstrated superior predictive performance compared to other models, including Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LGBM), Ridge Regression (Ridge), Least Absolute Shrinkage and Selection Operator (LASSO), and Support Vector Regression (SVR), specifically in predicting the maximum emission wavelength, the fluorescence quantum yield and stokes shift of multicolor CDs. The authors also implemented a computer algorithm for ranking importance, utilizing a method to calculate the significance of features. The outcomes revealed that the solvent was the primary factor influencing the maximum emission wavelength of multicolor CDs. The key determinant influencing the fluorescence quantum yield was identified to be the precursor ratio and the precursor type was

View Article Online
DOI: 10.1039/D4MA00505H

the main influencing factor of the stokes shift.

Xing et al. employed RF to facilitate the synthesis of CDs with predictable photoluminescence (PL).[122] In contrast to treating the precursors as constants, the variables in this context involve randomly chosen 202 combinations of precursors, specifically three-precursor combinations of 24 precursors. The wavelengths of the peaks with the strongest intensity and the longest wavelength under excitation wavelengths of 365 and 532 nm were used as output parameters. The other reaction parameters were fixed to 200°C and 10 h . RF model demonstrated the highest performance among the six models including KNN, AdaBoost, Bagging, DT, RF and SVM. It is shown that utilizing prediction data that encompasses the entire precursor combination space, the screening of CDs with specific PL wavelength features can be conducted much more effectively than through random trials.

He et al. established RF regression model for corrosion inhibitors based on hydrothermally synthesized CDs to predict the inhibition efficiency.[123] This model unveils the relationship between different synthesis parameters and the inhibition efficiency of the CDs. The dataset was created by combining 102 data points on CDs synthesis and inhibition efficiency, drawing from reported studies and the authors' own experimental findings. Typical input parameters such as CDs concentration in HCl, precursor type and quantity, solvent type and volume, and reaction time and temperature were selected. The inhibition efficiencies of CDs, calculated through potentiodynamic polarization (PDP), served as the output variable in the analysis. Utilizing the feature importance derived from the RF model, critical factors in the synthesis of CDs-based corrosion inhibitors were identified. The concentration of CDs in HCl emerges as the most influential factor affecting the inhibitory behaviors of the synthesized CDs, followed by N atomic content and reaction time. Additionally, the synthesis route is intelligently optimized using the Genetic Algorithm (GA), which is an optimization technique inspired by natural selection and genetics, utilizing a population-based approach with genetic operators to iteratively evolve solutions for a given problem. Successful controlled preparation of CDs-based corrosion inhibitors was achieved. By identifying and filtering

View Article Online
DOI: 10.1039/D4MA00505H

out unsatisfactory synthesis conditions, this approach significantly enhances the synthetic efficiency of CDs-based corrosion inhibitors (see Figure 5).

Other ML algorithms

Here we mention other ML algorithms used in CDs studies. Dager et al. detailed the production of mono-dispersed carbon quantum dots (C-QDs) through a single-step thermal decomposition procedure employing fennel seeds.[124] They employed ML techniques such as PCA (see[127] for more details on PCA), multivariate curve resolution (MCR),[128] and sparse non-negative matrix factorization (NMF)[129] to assess the PL of synthesized C-QDs with a focus on addressing two key questions: (i) the ability of ML to classify pH-dependent PL measurements, including spectra obtained at different pH levels and excitation wavelengths, and its capacity to suggest optimal excitation wavelengths for a comprehensive pH-dependent study; and (ii) whether ML can aid in identifying the source of the PL mechanism, considering that multiple PL measurements at varying pH levels and excitation wavelengths may activate different types of surface states. PL data were obtained through excitations at wavelengths of 200, 220, 240, 260, 280, 300, 320, and 340 nm, corresponding to pH values of 3, 5, 7, 9, 11, and 13. A total of forty-eight (48) PL measurements were conducted, each representing a single spectrum for 401 data points acquired in the spectral range 300-750 nm. PCA, MCR, and NMF were employed to identify the underlying mechanisms contributing to the PL behavior of the synthesized C-QDs.

Xu et al. used linear discriminant analysis (LDA)[130] and support vector machine (SVM)[131] to analyze multidimensional data of a CDs-based sensor array fabricated for the detection and differentiation of four tetracyclines (TC), including including tetracycline (TC), oxytetracycline (OTC), doxycycline (DOX), and metacycline (MTC).[125] A training data set comprising a matrix of 2 CDs, 4 TCs, and 5 replicates was created through the utilization of I/I_0 values. The reliability of the established fluorescence sensor array was confirmed by studying

View Article Online
DOI: 10.1039/D4MA00505H

52 unknown samples. At a concentration of 1.0 μM, four different TCs can be effectively clustered by SVM and LDA. Furthermore, the sensor array demonstrates the capability to effectively differentiate between individual TCs as well as binary mixtures of TCs and DOXs. The utilization of SVM presents an innovative option for array sensing systems in handling diverse data sets. The research illustrates the potential of the fluorescence sensor array in environmental monitoring and quantifying antibiotics (see Figure 6).

Summary and Future Perspectives

This comprehensive review explores the latest advancements in utilizing machine learning for CDs. We provide a concise summary of prevalent ML algorithms and examine recent research employing ML models for the prediction of CDs properties. ML models were employed to investigate the parameter space of CDs experiments and generate optimal input parameters for CDs. By leveraging the optimal parameters derived from ML for various CDs challenges, one can explore design strategies aimed at achieving high-performing CDs. While ML models are frequently perceived as "black box" models, the identified strategies can offer novel insights into enhancing the performance of CDs in various applications.

While artificial neural networks and gradient boosting algorithms have shown superior performance in several studies, research indicates that the optimal machine learning model can vary, even under identical input and target feature conditions. Hence, future research is needed to understand the performance, either theoretically or numerically, of various ML applications for CDs. Although achieving the true optimal experimental parameters remains a challenge in the field, there is optimism that ML will play a promising role in addressing this problem in the future. A promising avenue for enhancement involves establishing a more comprehensive models that incorporates both synthesis process-related and chemistry-related features.

The median of the sample size in the studies covered in this review is 357. For the

View Article Online
DOI: 10.1039/D4MA00505H

gradient boosting algorithms, this number is 467. To enhance the accuracy and applicability of ML approach, future endeavors should focus on collecting high-quality data for refining and updating the currently employed models. This continual improvement is crucial for advancing the development of more efficient CDs synthesis strategies.

Acknowledgement

ASJ thanks financial support of the King Fahd University of Petroleum and Minerals (KFUPM), the Deanship of Research Oversight and Coordination for funding this work through project INAM2206.

References

(1) Ghosal, K.; Ghosh, A. Carbon dots: The next generation platform for biomedical applications. *Mater Sci Eng C Mater Biol Appl* **2019**, *96*, 887–903.

(2) Sharma, A.; Das, J. Small molecules derived carbon dots: synthesis and applications in sensing, catalysis, imaging, and biomedicine. *Journal of Nanobiotechnology* **2019**, *17*, 92.

(3) Xiao, L.; Sun, H. Novel properties and applications of carbon nanodots. *Nanoscale Horizons* **2018**, *3*, 565–597.

(4) Ragazzon, G.; Cadranel, A.; Ushakova, E. V.; Wang, Y.; Guldi, D. M.; Rogach, A. L.; Kotov, N. A.; Prato, M. Optical processes in carbon nanocolloids. *Chem* **2021**, *7*, 606–628.

(5) Pan, L.; Sun, S.; Zhang, A.; Jiang, K.; Zhang, L.; Dong, C.; Huang, Q.; Wu, A.; Lin, H. Truly Fluorescent Excitation-Dependent Carbon Dots and Their Applications

View Article Online
DOI: 10.1039/D4MA00505H

in Multicolor Cellular Imaging and Multidimensional Sensing. *Advanced Materials* **2015**, *27*, 7782–7787.

(6) Benetti, D.; Jokar, E.; Yu, C.-H.; Fathi, A.; Zhao, H.; Vomiero, A.; Wei-Guang Diau, E.; Rosei, F. Hole-extraction and photostability enhancement in highly efficient inverted perovskite solar cells through carbon dot-based hybrid material. *Nano Energy* **2019**, *62*, 781–790.

(7) Wang, B.; Waterhouse, G. I. N.; Lu, S. Carbon dots: mysterious past, vibrant present, and expansive future. *Trends in Chemistry* **2023**, *5*, 76–87.

(8) Peng, J. et al. Graphene Quantum Dots Derived from Carbon Fibers. *Nano Letters* **2012**, *12*, 844–849.

(9) Rizzo, C.; Arcudi, F.; orević, L.; Dintcheva, N. T.; Noto, R.; D'Anna, F.; Prato, M. Nitrogen-Doped Carbon Nanodots-Ionogels: Preparation, Characterization, and Radical Scavenging Activity. *ACS Nano* **2018**, *12*, 1296–1305.

(10) Misra, S. K.; Srivastava, I.; Tripathi, I.; Daza, E.; Ostadhossein, F.; Pan, D. Macromolecularly "Caged" Carbon Nanoparticles for Intracellular Trafficking via Switchable Photoluminescence. *Journal of the American Chemical Society* **2017**, *139*, 1746–1749.

(11) Xu, X.; Zhang, K.; Zhao, L.; Li, C.; Bu, W.; Shen, Y.; Gu, Z.; Chang, B.; Zheng, C.; Lin, C.; Sun, H.; Yang, B. Aspirin-Based Carbon Dots, a Good Biocompatibility of Material Applied for Bioimaging and Anti-Inflammation. *ACS Applied Materials & Interfaces* **2016**, *8*, 32706–32716.

(12) Kim, T. H.; Sirdaarta, J. P.; Zhang, Q.; Eftekhari, E.; St. John, J.; Kennedy, D.; Cock, I. E.; Li, Q. Selective toxicity of hydroxyl-rich carbon nanodots for cancer research. *Nano Research* **2018**, *11*, 2204–2216.

View Article Online
DOI: 10.1039/D4MA00505H

(13) Li, S. et al. Targeted tumour theranostics in mice via carbon quantum dots structurally mimicking large amino acids. *Nature Biomedical Engineering* **2020**, *4*, 704–716.

(14) Mei, S.; Liu, X.; Zhang, W.; Liu, R.; Zheng, L.; Guo, R.; Tian, P. High-Bandwidth White-Light System Combining a Micro-LED with Perovskite Quantum Dots for Visible Light Communication. *ACS Applied Materials & Interfaces* **2018**, *10*, 5641–5648.

(15) Zhou, Z.; Tian, P.; Liu, X.; Mei, S.; Zhou, D.; Li, D.; Jing, P.; Zhang, W.; Guo, R.; Qu, S.; Rogach, A. L. Hydrogen Peroxide-Treated Carbon Dot Phosphor with a Bathochromic-Shifted, Aggregation-Enhanced Emission for Light-Emitting Devices and Visible Light Communication. *Advanced Science* **2018**, *5*, 1800369.

(16) Semeniuk, M.; Yi, Z.; Poursorkhabi, V.; Tjong, J.; Jaffer, S.; Lu, Z.-H.; Sain, M. Future Perspectives and Review on Organic Carbon Dots in Electronic Applications. *ACS Nano* **2019**, *13*, 6224–6255.

(17) Tetsuka, H.; Nagoya, A.; Fukusumi, T.; Matsui, T. Molecularly Designed, Nitrogen-Functionalized Graphene Quantum Dots for Optoelectronic Devices. *Advanced Materials* **2016**, *28*, 4632–4638.

(18) Arcudi, F.; Strauss, V.; orević, L.; Cadranel, A.; Guldi, D. M.; Prato, M. Porphyrin Antennas on Carbon Nanodots: Excited State Energy and Electron Transduction. *Angewandte Chemie International Edition* **2017**, *56*, 12097–12101.

(19) Yuan, T.; Meng, T.; He, P.; Shi, Y.; Li, Y.; Li, X.; Fan, L.; Yang, S. Carbon quantum dots: an emerging material for optoelectronic applications. *Journal of Materials Chemistry C* **2019**, *7*, 6820–6835.

(20) Li, Y.; Shu, H.; Niu, X.; Wang, J. Electronic and Optical Properties of Edge-Functionalized Graphene Quantum Dots and the Underlying Mechanism. *The Journal of Physical Chemistry C* **2015**, *119*, 24950–24957.

View Article Online
DOI: 10.1039/D4MA00505H

(21) Abdelsalam, H.; Elhaes, H.; Ibrahim, M. A. Tuning electronic properties in graphene quantum dots by chemical functionalization: Density functional theory calculations. *Chemical Physics Letters* **2018**, *695*, 138–148.

(22) Wang, B.; Yu, J.; Sui, L.; Zhu, S.; Tang, Z.; Yang, B.; Lu, S. Rational Design of Multi-Color-Emissive Carbon Dots in a Single Reaction System by Hydrothermal. *Advanced Science* **2021**, *8*, 2001453.

(23) Choi, Y.; Choi, Y.; Kwon, O.-H.; Kim, B.-S. Carbon Dots: Bottom-Up Syntheses, Properties, and Light-Harvesting Applications. *Chemistry –An Asian Journal* **2018**, *13*, 586–598.

(24) Dong, Y.; Chen, C.; Zheng, X.; Gao, L.; Cui, Z.; Yang, H.; Guo, C.; Chi, Y.; Li, C. M. One-step and high yield simultaneous preparation of single- and multi-layer graphene quantum dots from CX-72 carbon black. *Journal of Materials Chemistry* **2012**, *22*, 8764–8766.

(25) Li, Y.; Hu, Y.; Zhao, Y.; Shi, G.; Deng, L.; Hou, Y.; Qu, L. An Electrochemical Avenue to Green-Luminescent Graphene Quantum Dots as Potential Electron-Acceptors for Photovoltaics. *Advanced Materials* **2011**, *23*, 776–780.

(26) Sun, Y.-P. et al. Quantum-Sized Carbon Dots for Bright and Colorful Photoluminescence. *Journal of the American Chemical Society* **2006**, *128*, 7756–7757.

(27) Calabro, R. L.; Yang, D.-S.; Kim, D. Y. Liquid-phase laser ablation synthesis of graphene quantum dots from carbon nano-onions: Comparison with chemical oxidation. *Journal of Colloid and Interface Science* **2018**, *527*, 132–140.

(28) Xu, J.; Sahu, S.; Cao, L.; Anilkumar, P.; Tackett II, K. N.; Qian, H.; Bunker, C. E.; Guliants, E. A.; Parenzan, A.; Sun, Y.-P. Carbon Nanoparticles as Chromophores for Photon Harvesting and Photoconversion. *ChemPhysChem* **2011**, *12*, 3604–3608.

View Article Online
DOI: 10.1039/D4MA00505H

(29) Xu, X.; Ray, R.; Gu, Y.; Ploehn, H. J.; Gearheart, L.; Raker, K.; Scrivens, W. A. Electrophoretic Analysis and Purification of Fluorescent Single-Walled Carbon Nanotube Fragments. *Journal of the American Chemical Society* **2004**, *126*, 12736–12737.

(30) Li, X.; Wang, H.; Shimizu, Y.; Pyatenko, A.; Kawaguchi, K.; Koshizaki, N. Preparation of carbon quantum dots with tunable photoluminescence by rapid laser passivation in ordinary organic solvents. *Chemical Communications* **2011**, *47*, 932–934.

(31) Lee, J.; Kim, K.; Park, W. I.; Kim, B.-H.; Park, J. H.; Kim, T.-H.; Bong, S.; Kim, C.-H.; Chae, G.; Jun, M.; Hwang, Y.; Jung, Y. S.; Jeon, S. Uniform Graphene Quantum Dots Patterned from Self-Assembled Silica Nanodots. *Nano Letters* **2012**, *12*, 6078–6083.

(32) Ponomarenko, L. A.; Schedin, F.; Katsnelson, M. I.; Yang, R.; Hill, E. W.; Novoselov, K. S.; Geim, A. K. Chaotic Dirac Billiard in Graphene Quantum Dots. *Science* **2008**, *320*, 356–358.

(33) Weng, Y.; Li, Z.; Peng, L.; Zhang, W.; Chen, G. Fabrication of carbon quantum dots with nano-defined position and pattern in one step via sugar-electron-beam writing. *Nanoscale* **2017**, *9*, 19263–19270.

(34) Lim, S. Y.; Shen, W.; Gao, Z. Carbon quantum dots and their applications. *Chemical Society Reviews* **2015**, *44*, 362–381.

(35) de Medeiros, T. V.; Manioudakis, J.; Noun, F.; Macairan, J.-R.; Victoria, F.; Naccache, R. Microwave-assisted synthesis of carbon dots and their applications. *Journal of Materials Chemistry C* **2019**, *7*, 7175–7195.

(36) Arcudi, F.; orević, L.; Prato, M. Design, Synthesis, and Functionalization Strategies of Tailored Carbon Nanodots. *Accounts of Chemical Research* **2019**, *52*, 2070–2079.

View Article Online
DOI: 10.1039/D4MA00505H

(37) Suzuki, K.; Malfatti, L.; Takahashi, M.; Carboni, D.; Messina, F.; Tokudome, Y.; Takemoto, M.; Innocenzi, P. Design of Carbon Dots Photoluminescence through Organo-Functional Silane Grafting for Solid-State Emitting Devices. *Scientific Reports* **2017**, *7*, 5469.

(38) So, R. C.; Sanggo, J. E.; Jin, L.; Diaz, J. M. A.; Guerrero, R. A.; He, J. Gram-Scale Synthesis and Kinetic Study of Bright Carbon Dots from Citric Acid and Citrus japonica via a Microwave-Assisted Method. *ACS Omega* **2017**, *2*, 5196–5208.

(39) Khan, W. U.; Wang, D.; Zhang, W.; Tang, Z.; Ma, X.; Ding, X.; Du, S.; Wang, Y. High Quantum Yield Green-Emitting Carbon Dots for Fe() Detection, Biocompatible Fluorescent Ink and Cellular Imaging. *Scientific Reports* **2017**, *7*, 14866.

(40) Chen, B. B.; Liu, Z. X.; Deng, W. C.; Zhan, L.; Liu, M. L.; Huang, C. Z. A large-scale synthesis of photoluminescent carbon quantum dots: a self-exothermic reaction driving the formation of the nanocrystalline core at room temperature. *Green Chemistry* **2016**, *18*, 5127–5132.

(41) Li, T.; Shi, W.; E, S.; Mao, Q.; Chen, X. Green preparation of carbon dots with different surface states simultaneously at room temperature and their sensing applications. *Journal of Colloid and Interface Science* **2021**, *591*, 334–342.

(42) Branzi, L.; Lucchini, G.; Cattaruzza, E.; Pinna, N.; Benedetti, A.; Speghini, A. The formation mechanism and chirality evolution of chiral carbon dots prepared via radical assisted synthesis at room temperature. *Nanoscale* **2021**, *13*, 10478–10489.

(43) Zhao, P.; Li, X.; Baryshnikov, G.; Wu, B.; Ågren, H.; Zhang, J.; Zhu, L. One-step solvothermal synthesis of high-emissive amphiphilic carbon dots via rigidity derivation. *Chemical Science* **2018**, *9*, 1323–1329.

(44) Zhu, S.; Song, Y.; Zhao, X.; Shao, J.; Zhang, J.; Yang, B. The photoluminescence

View Article Online
DOI: 10.1039/D4MA00505H

mechanism in carbon dots (graphene quantum dots, carbon nanodots, and polymer dots): current state and future perspective. *Nano Research* **2015**, *8*, 355–381.

(45) Xia, C.; Zhu, S.; Feng, T.; Yang, M.; Yang, B. Evolution and Synthesis of Carbon Dots: From Carbon Dots to Carbonized Polymer Dots. *Advanced Science* **2019**, *6*, 1901316.

(46) Hola, K.; Zhang, Y.; Wang, Y.; Giannelis, E. P.; Zboril, R.; Rogach, A. L. Carbon dots—Emerging light emitters for bioimaging, cancer therapy and optoelectronics. *Nano Today* **2014**, *9*, 590–603.

(47) Lu, J.; Yang, J.-x.; Wang, J.; Lim, A.; Wang, S.; Loh, K. P. One-Pot Synthesis of Fluorescent Carbon Nanoribbons, Nanoparticles, and Graphene by the Exfoliation of Graphite in Ionic Liquids. *ACS Nano* **2009**, *3*, 2367–2375.

(48) Jin, S. H.; Kim, D. H.; Jun, G. H.; Hong, S. H.; Jeon, S. Tuning the Photoluminescence of Graphene Quantum Dots through the Charge Transfer Effect of Functional Groups. *ACS Nano* **2013**, *7*, 1239–1245.

(49) Sk, M. A.; Ananthanarayanan, A.; Huang, L.; Lim, K. H.; Chen, P. Revealing the tunable photoluminescence properties of graphene quantum dots. *Journal of Materials Chemistry C* **2014**, *2*, 6954–6960.

(50) Zhi, B. et al. Multicolor polymeric carbon dots: synthesis, separation and polyamide-supported molecular fluorescence. *Chemical Science* **2021**, *12*, 2441–2455.

(51) Fu, M.; Ehrat, F.; Wang, Y.; Milowska, K. Z.; Reckmeier, C.; Rogach, A. L.; Stolarczyk, J. K.; Urban, A. S.; Feldmann, J. Carbon Dots: A Unique Fluorescent Cocktail of Polycyclic Aromatic Hydrocarbons. *Nano Letters* **2015**, *15*, 6030–6035.

(52) Tepliakov, N. V.; Kundelev, E. V.; Khavlyuk, P. D.; Xiong, Y.; Leonov, M. Y.; Zhu, W.; Baranov, A. V.; Fedorov, A. V.; Rogach, A. L.; Rukhlenko, I. D. sp2–

View Article Online
DOI: 10.1039/D4MA00505H

sp3-Hybridized Atomic Domains Determine Optical Features of Carbon Dots. *ACS Nano* **2019**, *13*, 10737–10744.

(53) Meiling, T. T.; Schürmann, R.; Vogel, S.; Ebel, K.; Nicolas, C.; Milosavljević, A. R.; Bald, I. Photophysics and Chemistry of Nitrogen-Doped Carbon Nanodots with High Photoluminescence Quantum Yield. *The Journal of Physical Chemistry C* **2018**, *122*, 10217–10230.

(54) Pires, N. R.; Santos, C. M. W.; Sousa, R. R.; Paula, R. C. M. d.; Cunha, P. L. R.; Feitosa, J. P. A. Novel and Fast Microwave-Assisted Synthesis of Carbon Quantum Dots from Raw Cashew Gum. *Journal of the Brazilian Chemical Society* **2015**, *26*.

(55) Parvin, N.; Mandal, T. K. Synthesis of a highly fluorescence nitrogen-doped carbon quantum dots bioimaging probe and its in vivo clearance and printing applications. *RSC Advances* **2016**, *6*, 18134–18140.

(56) Zhang, J.; Chen, X.; Li, Y.; Han, S.; Du, Y.; Liu, H. A nitrogen doped carbon quantum dot-enhanced chemiluminescence method for the determination of Mn2+. *Analytical Methods* **2018**, *10*, 541–547.

(57) Liu, X.; Pang, J.; Xu, F.; Zhang, X. Simple Approach to Synthesize Amino-Functionalized Carbon Dots by Carbonization of Chitosan. *Scientific Reports* **2016**, *6*, 31100.

(58) Dong, Y.; Wan, L.; Cai, J.; Fang, Q.; Chi, Y.; Chen, G. Natural carbon-based dots from humic substances. *Scientific Reports* **2015**, *5*, 10037.

(59) Xue, M.; Zhan, Z.; Zou, M.; Zhang, L.; Zhao, S. Green synthesis of stable and biocompatible fluorescent carbon dots from peanut shells for multicolor living cell imaging. *New Journal of Chemistry* **2016**, *40*, 1698–1703.

View Article Online
DOI: 10.1039/D4MA00505H

(60) Guo, X.; Zhang, H.; Sun, H.; Tade, M. O.; Wang, S. Green Synthesis of Carbon Quantum Dots for Sensitized Solar Cells. *ChemPhotoChem* **2017**, *1*, 116–119.

(61) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **1964**, *136*, B864–B871.

(62) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, *140*, A1133–A1138.

(63) Dral, P. O.; Clark, T. Semiempirical UNO–CAS and UNO–CI: Method and Applications in Nanoelectronics. *The Journal of Physical Chemistry A* **2011**, *115*, 11303–11312.

(64) Vorontsov, A. V.; Tretyakov, E. V. Determination of graphene's edge energy using hexagonal graphene quantum dots and PM7 method. *Physical Chemistry Chemical Physics* **2018**, *20*, 14740–14752.

(65) Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *Journal of Molecular Modeling* **2013**, *19*, 1–32.

(66) Budyka, M. F. Semiempirical study on the absorption spectra of the coronene-like molecular models of graphene quantum dots. *Spectrochim Acta A Mol Biomol Spectrosc* **2019**, *207*, 1–5.

(67) Margraf, J. T.; Strauss, V.; Guldi, D. M.; Clark, T. The Electronic Structure of Amorphous Carbon Nanodots. *The Journal of Physical Chemistry B* **2015**, *119*, 7258–7265.

(68) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260–7268.

View Article Online
DOI: 10.1039/D4MA00505H

(69) Seifert, G. Tight-Binding Density Functional Theory: An Approximate KohnSham DFT Scheme. *The Journal of Physical Chemistry A* **2007**, *111*, 5609–5613.

(70) Seifert, G.; Porezag, D.; Frauenheim, T. Calculations of molecules, clusters, and solids with a simplified LCAO-DFT-LDA scheme. **1996**, *58*, 185–192.

(71) Oliveira, A. F.; Seifert, G.; Heine, T.; Duarte, H. A. Density-functional based tight-binding: an approximate DFT method. *Journal of the Brazilian Chemical Society* **2009**, *20*.

(72) Page, A. J.; Yamane, H.; Ohta, Y.; Irle, S.; Morokuma, K. QM/MD Simulation of SWNT Nucleation on Transition-Metal Carbide Nanoparticles. *Journal of the American Chemical Society* **2010**, *132*, 15699–15707.

(73) Lei, T.; Guo, W.; Liu, Q.; Jiao, H.; Cao, D.-B.; Teng, B.; Li, Y.-W.; Liu, X.; Wen, X.-D. Mechanism of Graphene Formation via Detonation Synthesis: A DFTB Nanoreactor Approach. *Journal of Chemical Theory and Computation* **2019**, *15*, 3654–3665.

(74) Botu, V.; Ramprasad, R. Adaptive machine learning framework to accelerate ab initio molecular dynamics. *International Journal of Quantum Chemistry* **2015**, *115*, 1074–1083.

(75) Mitchell, T. M. *Machine learning*; McGraw-hill New York, 1997; Vol. 1.

(76) Koh, D.-M.; Papanikolaou, N.; Bick, U.; Illing, R.; Kahn, C. E.; Kalpathi-Cramer, J.; Matos, C.; Martí-Bonmatí, L.; Miles, A.; Mun, S. K.; Napel, S.; Rockall, A.; Sala, E.; Strickland, N.; Prior, F. Artificial intelligence and machine learning in cancer imaging. *Communications Medicine* **2022**, *2*, 133.

(77) Cambria, E.; White, B. Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]. *Computational Intelligence Magazine, IEEE* **2014**, *9*, 48–57.

View Article Online
DOI: 10.1039/D4MA00505H

(78) Jin, C.; Yu, H.; Ke, J.; Ding, P.; Yi, Y.; Jiang, X.; Duan, X.; Tang, J.; Chang, D. T.; Wu, X.; Gao, F.; Li, R. Predicting treatment response from longitudinal images using multi-task deep learning. *Nature Communications* **2021**, *12*, 1851.

(79) Kalasin, S.; Sangnuang, P.; Surareungchai, W. Lab-on-Eyeglasses to Monitor Kidneys and Strengthen Vulnerable Populations in Pandemics: Machine Learning in Predicting Serum Creatinine Using Tear Creatinine. *Analytical Chemistry* **2021**, *93*, 10661–10671.

(80) Kiyasseh, D.; Zhu, T.; Clifton, D. A clinical deep learning framework for continually learning from cardiac signals across diseases, time, modalities, and institutions. *Nature Communications* **2021**, *12*, 4221.

(81) Xie, Y.; Zhang, C.; Hu, X.; Zhang, C.; Kelley, S. P.; Atwood, J. L.; Lin, J. Machine Learning Assisted Synthesis of Metal–Organic Nanocapsules. *Journal of the American Chemical Society* **2020**, *142*, 1475–1481.

(82) Han, Y.; Tang, B.; Wang, L.; Bao, H.; Lu, Y.; Guan, C.; Zhang, L.; Le, M.; Liu, Z.; Wu, M. Machine-Learning-Driven Synthesis of Carbon Dots with Enhanced Quantum Yields. *ACS Nano* **2020**, *14*, 14761–14768.

(83) Wang, X.; Wang, B.; Wang, H.; Zhang, T.; Qi, H.; Wu, Z.; Ma, Y.; Huang, H.; Shao, M.; Liu, Y.; Li, Y.; Kang, Z. Carbon-Dot-Based White-Light-Emitting Diodes with Adjustable Correlated Color Temperature Guided by Machine Learning. *Angewandte Chemie International Edition* **2021**, *60*, 12585–12590.

(84) Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710.

(85) Im, J.; Lee, S.; Ko, T.-W.; Kim, H. W.; Hyon, Y.; Chang, H. Identifying Pb-free perovskites for solar cells by machine learning. *npj Computational Materials* **2019**, *5*, 37.

View Article Online
DOI: 10.1039/D4MA00505H

(86) Jin, H.; Zhang, H.; Li, J.; Wang, T.; Wan, L.; Guo, H.; Wei, Y. Discovery of Novel Two-Dimensional Photovoltaic Materials Accelerated by Machine Learning. *The Journal of Physical Chemistry Letters* **2020**, *11*, 3075–3081.

(87) Pradhan, D. K.; Kumari, S.; Strelcov, E.; Pradhan, D. K.; Katiyar, R. S.; Kalinin, S. V.; Laanait, N.; Vasudevan, R. K. Reconstructing phase diagrams from local measurements via Gaussian processes: mapping the temperature-composition space to confidence. *npj Computational Materials* **2018**, *4*, 23.

(88) Timoshenko, J.; Lu, D.; Lin, Y.; Frenkel, A. I. Supervised Machine-Learning-Based Determination of Three-Dimensional Structure of Metallic Nanoparticles. *The Journal of Physical Chemistry Letters* **2017**, *8*, 5091–5098.

(89) Takahashi, K.; Takahashi, L. Creating Machine Learning-Driven Material Recipes Based on Crystal Structure. *The Journal of Physical Chemistry Letters* **2019**, *10*, 283–288.

(90) Zhou, Q.; Lu, S.; Wu, Y.; Wang, J. Property-Oriented Material Design Based on a Data-Driven Machine Learning Technique. *The Journal of Physical Chemistry Letters* **2020**, *11*, 3920–3927.

(91) Sahu, H.; Ma, H. Unraveling Correlations between Molecular Properties and Device Parameters of Organic Solar Cells Using Machine Learning. *The Journal of Physical Chemistry Letters* **2019**, *10*, 7277–7284.

(92) Li, D.; Li, X.; Zhang, Y.; Sun, L.; Yuan, S. Four Methods to Estimate Minimum Miscibility Pressure of CO2-Oil Based on Machine Learning. *Chinese Journal of Chemistry* **2019**, *37*, 1271–1278.

(93) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361*, 360–365.

View Article Online
DOI: 10.1039/D4MA00505H

(94) Noh, J.; Kim, J.; Stein, H. S.; Sanchez-Lengeling, B.; Gregoire, J. M.; Aspuru-Guzik, A.; Jung, Y. Inverse Design of Solid-State Materials via a Continuous Representation. *Matter* **2019**, *1*, 1370–1384.

(95) Yao, Z.; Sánchez-Lengeling, B.; Bobbitt, N. S.; Bucior, B. J.; Kumar, S. G. H.; Collins, S. P.; Burns, T.; Woo, T. K.; Farha, O. K.; Snurr, R. Q.; Aspuru-Guzik, A. Inverse design of nanoporous crystalline reticular materials with deep generative models. *Nature Machine Intelligence* **2021**, *3*, 76–86.

(96) Ma, Y.; Jin, T.; Choudhury, R.; Cheng, Q.; Miao, Y.; Zheng, C.; Min, W.; Yang, Y. Understanding the Correlation between Lithium Dendrite Growth and Local Material Properties by Machine Learning. **2021**, *168*, 090523.

(97) Zhu, J.; Ren, Z.; Lee, C. Toward Healthcare Diagnoses by Machine-Learning-Enabled Volatile Organic Compound Identification. *ACS Nano* **2021**, *15*, 894–903.

(98) Tao, H.; Wu, T.; Aldeghi, M.; Wu, T. C.; Aspuru-Guzik, A.; Kumacheva, E. Nanoparticle synthesis assisted by machine learning. *Nature Reviews Materials* **2021**, *6*, 701–716.

(99) Bartolomei, B.; Dosso, J.; Prato, M. New trends in nonconventional carbon dot synthesis. *Trends in Chemistry* **2021**, *3*, 943–953.

(100) Chen, C.-T.; Gu, G. X. Machine learning for composite materials. **2019**, *9*, 556–566.

(101) Huang, M.; Li, Z.; Zhu, H. Recent Advances of Graphene and Related Materials in Artificial Intelligence. *Advanced Intelligent Systems* **2022**, *4*, 2200077.

(102) Langer, M.; Paloncýová, M.; Medveď, M.; Pykal, M.; Nachtigallová, D.; Shi, B.; Aquino, A. J. A.; Lischka, H.; Otyepka, M. Progress and challenges in understanding of photoluminescence properties of carbon dots based on theoretical computations. *Applied Materials Today* **2021**, *22*, 100924.

View Article Online
DOI: 10.1039/D4MA00505H

(103) Munyebvu, N.; Lane, E.; Grisan, E.; Howes, P. D. Accelerating colloidal quantum dot innovation with algorithms and automation. *Materials Advances* **2022**, *3*, 6950–6967.

(104) Peng, J.; Muhammad, R.; Wang, S.-L.; Zhong, H.-Z. How Machine Learning Accelerates the Development of Quantum Dots?†. *Chinese Journal of Chemistry* **2021**, *39*, 181–188.

(105) Tamtaji, M.; Tyagi, A.; You, C. Y.; Galligan, P. R.; Liu, H.; Liu, Z.; Karimi, R.; Cai, Y.; Roxas, A. P.; Wong, H.; Luo, Z. Singlet Oxygen Photosensitization Using Graphene-Based Structures and Immobilized Dyes: A Review. *ACS Applied Nano Materials* **2021**, *4*, 7563–7586.

(106) Zhu, M.-X.; Deng, T.; Dong, L.; Chen, J.-M.; Dang, Z.-M. Review of machine learning-driven design of polymer-based dielectrics. *IET Nanodielectrics* **2022**, *5*, 24–38.

(107) Armida, S. A.; Ebrahimibagha, D.; Ray, M.; Datta, S. Assessing thermoelectric performance of quasi 0D carbon and polyaniline nanocomposites using machine learning. *Advanced Composite Materials* 1–23.

(108) Zhang, Q.; Tao, Y.; Tang, B.; Yang, J.; Liang, H.; Wang, B.; Wang, J.; Jiang, X.; Ji, L.; Li, S. Graphene Quantum Dots with Improved Fluorescence Activity via Machine Learning: Implications for Fluorescence Monitoring. *ACS Applied Nano Materials* **2022**, *5*, 2728–2737.

(109) Tuchin, V. S.; Stepanidenko, E. A.; Vedernikova, A. A.; Cherevkov, S. A.; Li, D.; Li, L.; Döring, A.; Otyepka, M.; Ushakova, E. V.; Rogach, A. L. Optical Properties Prediction for Red and Near-Infrared Emitting Carbon Dots Using Machine Learning. *Small* **2024**, *n/a*, 2310402.

(110) Döring, A.; Qiu, Y.; Rogach, A. L. Improving the Accuracy of Carbon Dot Temperature Sensing Using Multi-Dimensional Machine Learning. *ACS Applied Nano Materials* **2024**, *7*, 2258–2269.

View Article Online
DOI: 10.1039/D4MA00505H

(111) Chen, J.; Luo, J. B.; Hu, M. Y.; Zhou, J.; Huang, C. Z.; Liu, H. Controlled Synthesis of Multicolor Carbon Dots Assisted by Machine Learning. *Advanced Functional Materials* **2023**, *33*, 2210095.

(112) Pandit, S.; Banerjee, T.; Srivastava, I.; Nie, S.; Pan, D. Machine Learning-Assisted Array-Based Biomolecular Sensing Using Surface-Functionalized Carbon Dots. *ACS Sensors* **2019**, *4*, 2730–2737.

(113) Wang, X.-Y.; Chen, B.-B.; Zhang, J.; Zhou, Z.-R.; Lv, J.; Geng, X.-P.; Qian, R.-C. Exploiting deep learning for predictable carbon dot design. *Chem. Commun.* **2021**, *57*, 532–535.

(114) Senanayake, R. D.; Yao, X.; Froehlich, C. E.; Cahill, M. S.; Sheldon, T. R.; McIntire, M.; Haynes, C. L.; Hernandez, R. Machine Learning-Assisted Carbon Dot Synthesis: Prediction of Emission Color and Wavelength. *Journal of Chemical Information and Modeling* **2022**, *62*, 5918–5928.

(115) Tuccitto, N.; Fichera, L.; Ruffino, R.; Cantaro, V.; Sfuncia, G.; Nicotra, G.; Sfrazzetto, G. T.; Li-Destri, G.; Valenti, A.; Licciardello, A.; Torrisi, A. Carbon Quantum Dots as Fluorescence Nanochemosensors for Selective Detection of Amino Acids. *ACS Applied Nano Materials* **2021**, *4*, 6250–6256.

(116) Yahaya Pudza, M.; Zainal Abidin, Z.; Abdul Rashid, S.; Md Yasin, F.; Noor, A. S.; Issa, M. A. Sustainable Synthesis Processes for Carbon Dots through Response Surface Methodology and Artificial Neural Network. **2019**, *7*.

(117) Döring, A.; Rogach, A. L. Utilizing Deep Learning to Enhance Optical Sensing of Ethanol Content Based on Luminescent Carbon Dots. *ACS Applied Nano Materials* **2022**, *5*, 11208–11218.

(118) Tang, B.; Lu, Y.; Zhou, J.; Chouhan, T.; Wang, H.; Golani, P.; Xu, M.; Xu, Q.;

View Article Online
DOI: 10.1039/D4MA00505H

Guan, C.; Liu, Z. Machine learning-guided synthesis of advanced inorganic materials. *Materials Today* **2020**, *41*, 72–80.

(119) Hong, Q.; Wang, X.-Y.; Gao, Y.-T.; Lv, J.; Chen, B.-B.; Li, D.-W.; Qian, R.-C. Customized Carbon Dots with Predictable Optical Properties Synthesized at Room Temperature Guided by Machine Learning. *Chemistry of Materials* **2022**, *34*, 998–1009.

(120) Wang, X.; Chen, S.; Ma, Y.; Zhang, T.; Zhao, Y.; He, T.; Huang, H.; Zhang, S.; Rong, J.; Shi, C.; Tang, K.; Liu, Y.; Kang, Z. Continuous Homogeneous Catalytic Oxidation of C–H Bonds by Metal-Free Carbon Dots with a Poly(ascorbic acid) Structure. *ACS Applied Materials & Interfaces* **2022**, *14*, 26682–26689.

(121) Chen, J.; Zhang, M.; Xu, Z.; Ma, R.; Shi, Q. Machine-learning analysis to predict the fluorescence quantum yield of carbon quantum dots in biochar. *Science of The Total Environment* **2023**, *896*, 165136.

(122) Xing, C.; Chen, G.; Zhu, X.; An, J.; Bao, J.; Wang, X.; Zhou, X.; Du, X.; Xu, X. Synthesis of carbon dots with predictable photoluminescence by the aid of machine learning. *Nano Research* **2023**,

(123) He, H.; E, S.; Ai, L.; Wang, X.; Yao, J.; He, C.; Cheng, B. Exploiting machine learning for controlled synthesis of carbon dots-based corrosion inhibitors. *Journal of Cleaner Production* **2023**, *419*, 138210.

(124) Dager, A.; Uchida, T.; Maekawa, T.; Tachibana, M. Synthesis and characterization of Mono-disperse Carbon Quantum Dots from Fennel Seeds: Photoluminescence analysis using Machine Learning. *Scientific Reports* **2019**, *9*, 14004.

(125) Xu, Z.; Wang, Z.; Liu, M.; Yan, B.; Ren, X.; Gao, Z. Machine learning assisted dual-channel carbon quantum dots-based fluorescence sensor array for detection of

View Article Online
DOI: 10.1039/D4MA00505H

tetracyclines. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2020**, *232*, 118147.

(126) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA, 2016; pp 785–794.

(127) Maslova, O. A.; Guimbretière, G.; Ammar, M. R.; Desgranges, L.; Jégou, C.; Canizarès, A.; Simon, P. Raman imaging and principal component analysis-based data processing on uranium oxide ceramics. *Materials Characterization* **2017**, *129*, 260–269.

(128) Felten, J.; Hall, H.; Jaumot, J.; Tauler, R.; de Juan, A.; Gorzsás, A. Vibrational spectroscopic image analysis of biological material using multivariate curve resolution–alternating least squares (MCR-ALS). *Nature Protocols* **2015**, *10*, 217–240.

(129) Shiga, M.; Tatsumi, K.; Muto, S.; Tsuda, K.; Yamamoto, Y.; Mori, T.; Tanji, T. Sparse modeling of EELS and EDX spectral imaging data by nonnegative matrix factorization. *Ultramicroscopy* **2016**, *170*, 43–59.

(130) Yan, P.; Li, X.; Dong, Y.; Li, B.; Wu, Y. A pH-based sensor array for the detection and identification of proteins using CdSe/ZnS quantum dots as an indicator. *Analyst* **2019**, *144*, 2891–2897.

(131) Li, L.-L.; Zhao, X.; Tseng, M.-L.; Tan, R. R. Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. *Journal of Cleaner Production* **2020**, *242*, 118447.

View Article Online
DOI: 10.1039/D4MA00505H

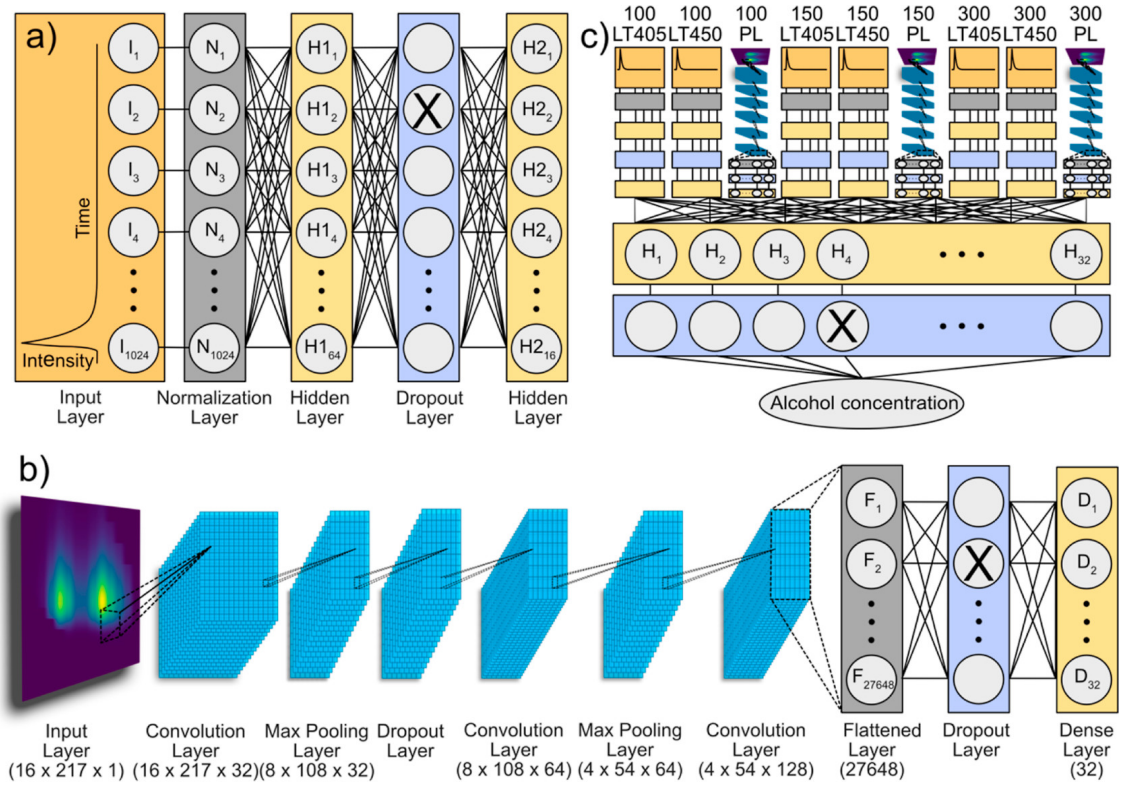

Figure 2: Multi-channel deep learning model: (a) structure of the PL decay channel. The input layer takes 1024 intensity integers as input. After normalization, data are passed through a dense layer (64 neurons), a dropout layer (dropout = 0.01), and a second dense layer (16 neurons). (b) Structure of the PL map channel. The input layer takes a $16 \times 217 \times 1$ matrix as input. It is passed through a series of convolution, maximum pooling, and dropout layers before it is flattened and fed through another dropout layer and dense layer (32 neurons). (c) Example of a multi-channel model with 9 inputs. The respective input data are passed through either a PL decay channel or a PL map channel. Those channels are concatenated before being passed through a dense layer (32 neurons) and a dropout layer (dropout = 0.3) to predict the ethanol concentration as the target variable.Reprinted with permission from.[117] Copyright 2022 American Chemical Society.

View Article Online
DOI: 10.1039/D4MA00505H

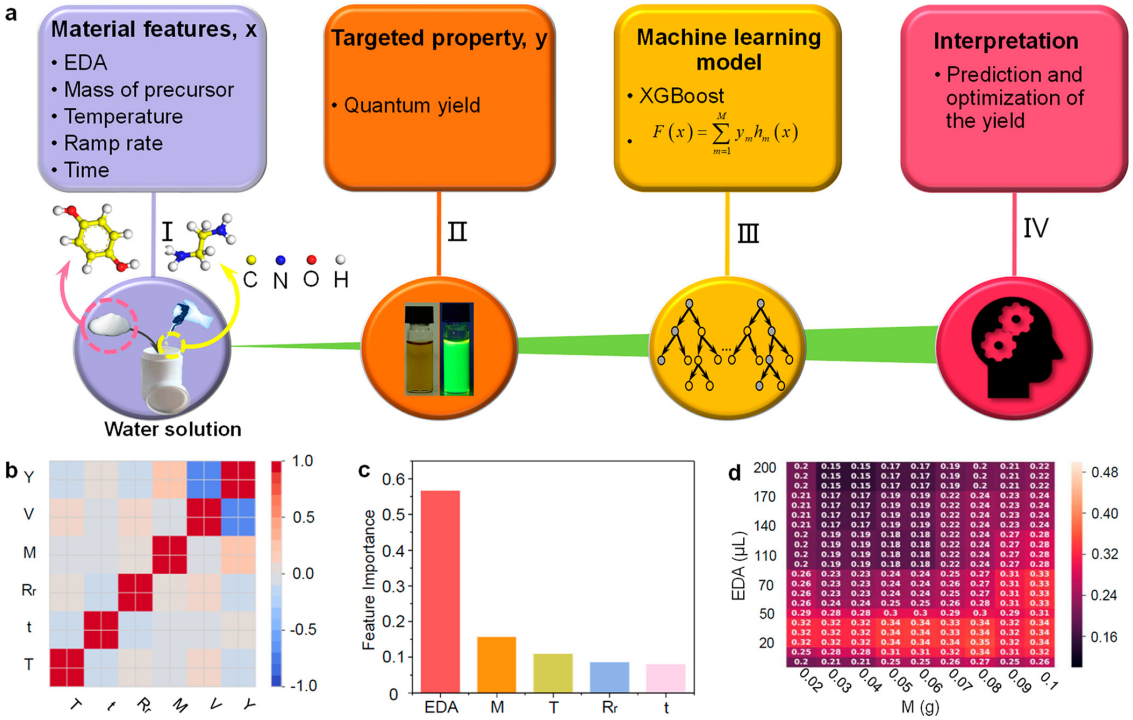

Figure 3: Application of ML for guided synthesis of CDs. (a) Design framework for the guided synthesis of CDs with a large QY based on ML and hydrothermal experiments. (b) The heat map of the Pearson's correlation coefficient matrix among the selected features of hydrothermal-grown CDs. (c) Feature importance retrieved from XGBoost-R that learns from the full data set. The most important features are EDA and M. (d) Predictions from the trained model, which is represented by the matrix formed by the two most important features. Reprinted with permission from.[82] Copyright 2020 American Chemical Society.

View Article Online
DOI: 10.1039/D4MA00505H

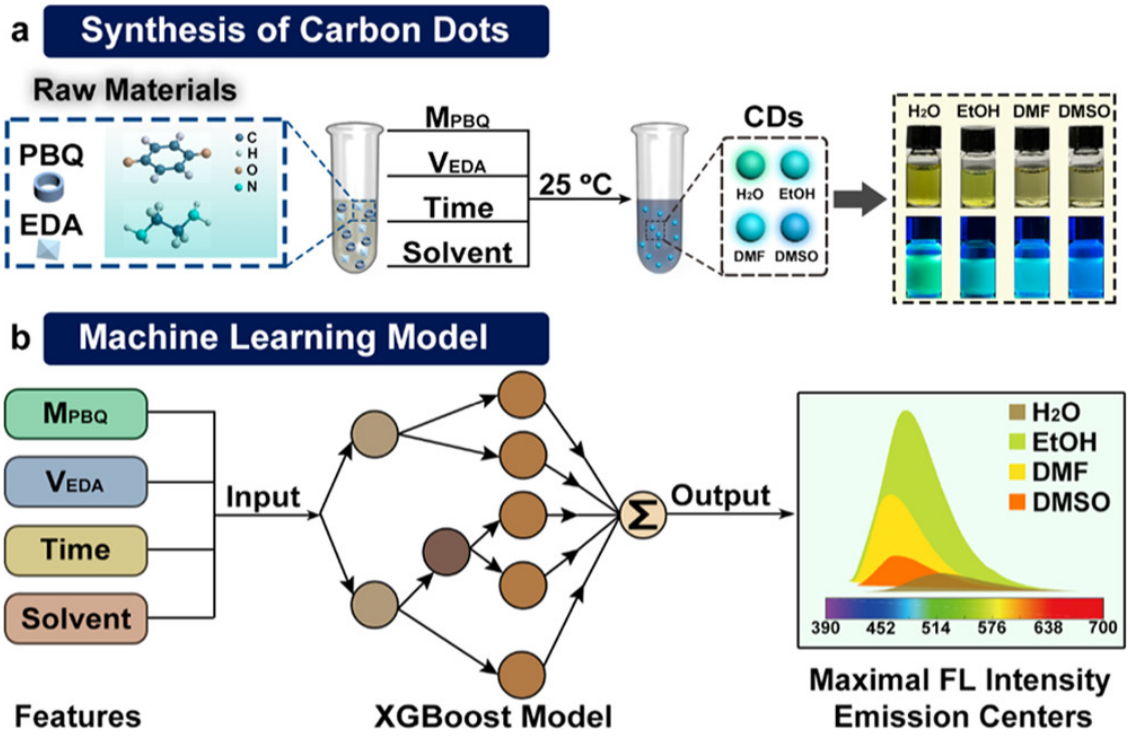

Figure 4: Schematic Illustration of Machine Learning Guiding the Synthesis of CDs.(a) Synthetic process of CDs. (b) Prediction of CD optical properties using machine learning models. Reprinted with permission from.[119] Copyright 2022 American Chemical Society.

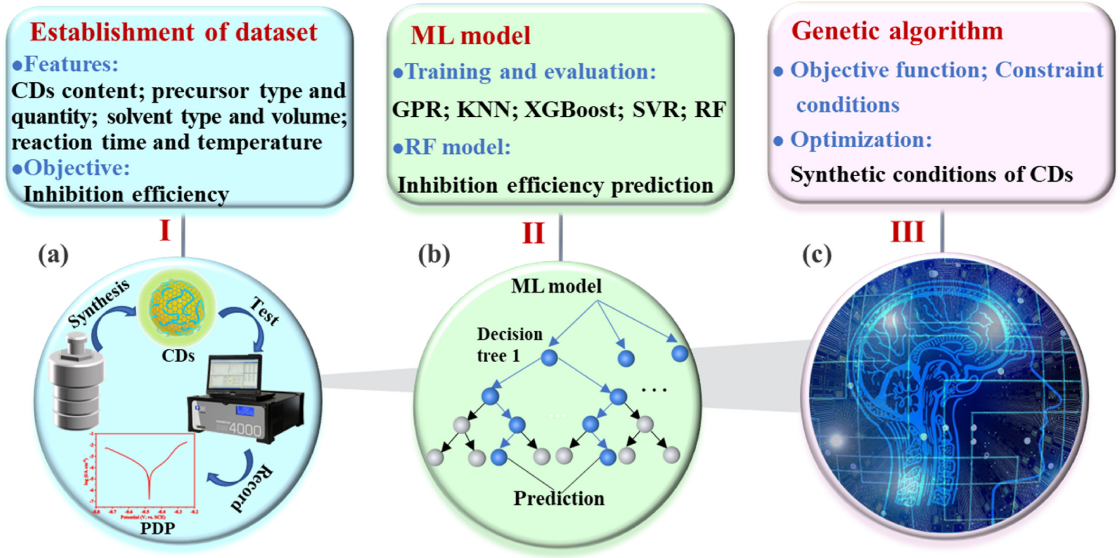

Figure 5: Application of ML for controlled synthesis of CDs-based corrosion inhibitors: (a) establishment of the dataset; (b) modelling for inhibition efficiency prediction; (c) synthetic optimization of CDs.Reprinted with permission from.[123] Copyright 2023 Elsevier.

View Article Online
DOI: 10.1039/D4MA00505H

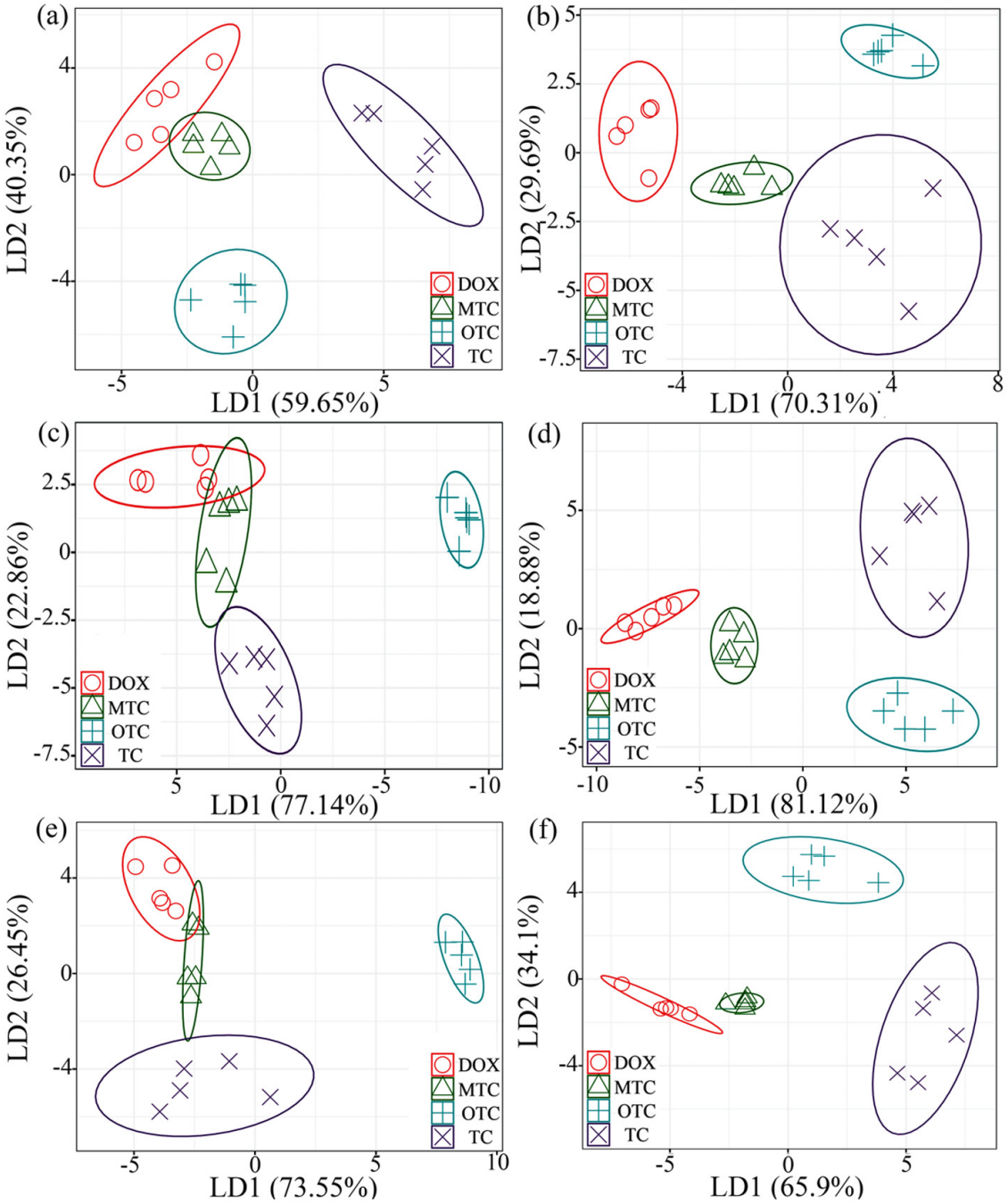

Figure 6: Two-dimensional LDA score plot of the fluorescence sensor array for the discrimination of the four TCs at different concentrations: (a) $1.0\mu M$; (b) $10\mu M$; (c) $25\mu M$; (d) $50\mu M$; (e)$100\mu M$; (f)$150\mu M$. (QR-CDs, $13.3\mu g\ mL^{-1}$; CPC-CDs, $60\mu g\ mL^{-1}$.).Reprinted with permission from.[125]Copyright 2020 Elsevier.

Data availability statement

This is a review article. No data is used to produce in the process.

View Article Online
DOI: 10.1039/D4MA00505H