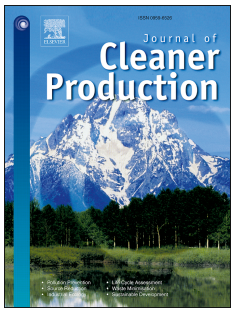


Exploiting machine learning for controlled synthesis of carbon dots-based corrosion inhibitors

Haijie He, Shuang E, Li Ai, Xiaogang Wang, Jun Yao, Chuang He, Boyuan Cheng

PII: S0959-6526(23)02368-5

DOI: https://doi.org/10.1016/j.jclepro.2023.138210

Reference: JCLP 138210

To appear in: *Journal of Cleaner Production*

Received Date: 25 May 2023

Revised Date: 19 July 2023

Accepted Date: 20 July 2023

Please cite this article as: 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), doi: https://doi.org/10.1016/j.jclepro.2023.138210.

Credit Author Statement

Haijie He: data processing, writing original draft preparation. **Shuang E**: data processing, writing original draft preparation. **Li Ai**: data collection, data processing. **Xiaogang Wang**: data collection. **Jun Yao**: manuscript reviewing. **Chuang He**: methodology, project administration, funding acquisition. **Boyuan Cheng**: methodology, manuscript reviewing, modeling.

Exploiting machine learning for controlled synthesis of carbon dots-based corrosion inhibitors

Haijie He[&a], Shuang E[&b], Li Ai[c], Xiaogang Wang[a], Jun Yao[a], Chuang He[1*a], Boyuan Cheng[*d]

[a] School of Civil Engineering and Architecture, Taizhou University, Taizhou, Zhejiang, PR China

[b] College of Life Science, Dalian Minzu University, Dalian, Liaoning, PR China

[c] Department of Civil and Environmental Engineering, University of South Carolina, Columbia, SC, USA

[d] College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, Guangdong, PR China

Corresponding author:

Dr. Chuang He

Mailing Address: School of Civil Engineering and Architecture, Taizhou University, Taizhou 318000, Zhejiang, PR China

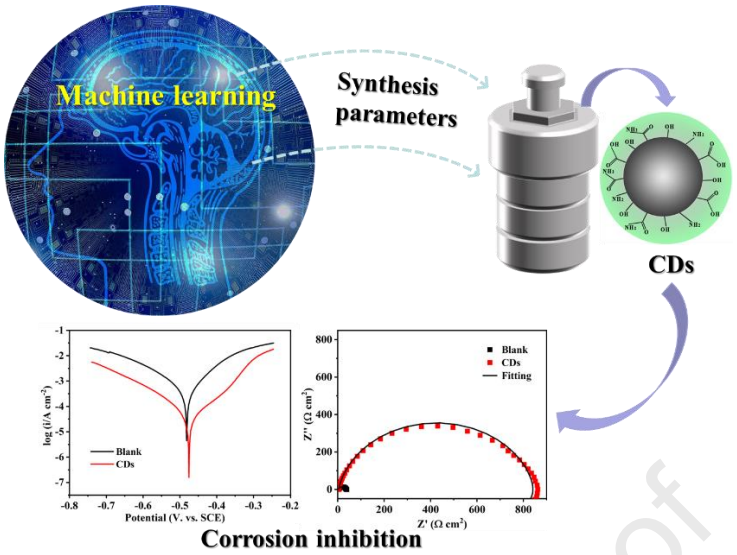
E-mail addresses: 584650078@qq.com; hechuang@szu.edu.cn

Dr. Boyuan Cheng

Mailing Address: College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518060, Guangdong, P.R. China

E-mail addresses: chengboyuan2020@email.szu.edu.cn

Declarations of Interest: None



Exploiting machine learning for controlled synthesis of carbon dots-based corrosion inhibitors

Haijie He[&a], Shuang E[&b], Li Ai[c], Xiaogang Wang[a], Jun Yao[a], Chuang He[*a], Boyuan Cheng[*d]

[a] School of Civil Engineering and Architecture, Taizhou University, Taizhou, Zhejiang, PR China

[b] College of Life Science, Dalian Minzu University, Dalian, Liaoning, PR China

[c] Department of Civil and Environmental Engineering, University of South Carolina, Columbia, SC, USA

[d] College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, Guangdong, PR China

Abstract

Benefitting from their prominent corrosion inhibition properties, excellent water solubility and benign environmental friendliness, carbon dots (CDs) have functioned as an ideal candidate for next-generation green corrosion inhibitors. However, the extensive adoption of the trial-and-error route driven by artificial experience in the preparation of CDs-based corrosion inhibitors leads to resource waste and environmental implications, detrimental to their sustainable development. It is still a considerable challenge to controllably prepare CDs-based corrosion inhibitors with the predictable inhibition efficiency. Herein, firstly exploiting a data-driven machine learning (ML) approach, this study aims to precisely predict the inhibition efficiency of CDs and optimize their synthesis route, resulting in the controlled synthesis of CDs-based corrosion inhibitors. Specifically, the dataset is constructed by combining 102 data points on CDs synthesis and inhibition efficiency from numerous published studies and our own experiments. After training and evaluation of different ML models, the Random Forest (RF) ML regression model is chosen with the lowest root-mean-

[&] These authors contributed equally to this work and should be considered co-first authors.

[*] Corresponding author.

E-mail: 584650078@qq.com; hechuang@szu.edu.cn (C. He).

[*] Corresponding author.

E-mail: chengboyuan2020@email.szu.edu.cn (B.Y. Cheng).

square error and mean absolute error as well as the highest coefficient of determination. The results show that this RF model can comprehensively reveal the relationship between various hydrothermal synthesis parameters and the inhibition efficiency. Guided by the RF model, the inhibition efficiencies of CDs-based corrosion inhibitors are accurately predicted with an error less than 10%, and based on the genetic algorithm, their synthesis route is intelligently optimized. This work demonstrates the feasibility of ML techniques in guiding the optimization of synthesis conditions for CDs-based corrosion inhibitors. This optimization process results in reduced development time and cost, contributing to the sustainability and cleaner production of inhibitors.

Key Words: Carbon dots; Corrosion inhibitor; Machine learning; Inhibition efficiency.

1. Introduction

Metal corrosion severely threatens the safety of metallic equipment, and critically aggravates environmental pollution and economic losses, thus always serving as one of the most concerned topics of mankind (Long et al., 2022b; Madlangbayan et al., 2021). How to effectively protect metal from corrosion is always a typical hot and difficult problem (Astuti et al., 2022). As a cost-effective, simple and high-efficiency approach, corrosion inhibitors have attracted significant attention and extensively employed to mitigate the metallic corrosion rate. Conventional corrosion inhibitors containing toxic elements tend to damage human health and pollute the environment (He et al., 2022a). Hence, massive efforts have been made to develop environmentally-friendly green corrosion inhibitors (He et al., 2022a; Kobzar and Fatyeyeva, 2021; Tan et al., 2021). Among them, carbon dots (CDs) are considered as one of the most promising candidates for next-generation green corrosion inhibitors, owing to their prominent corrosion inhibition properties, excellent water solubility, benign environmental friendliness, high chemical inertness, easily available precursors and various synthetic methods (Cui et al., 2017; He et al., 2022a; Li and Gong, 2022; Li et al., 2023; Long et al., 2022a; Long et al., 2022b; Qiang et al., 2019; Ren et al., 2022; Ye, Yuwei et al., 2020; Ye, Y. et al., 2020). Up to now, CDs have demonstrated remarkable inhibition effects for various metals in acid/salt corrosive solution (Cui et al., 2017; He et al., 2022a; Long et al., 2022a; Long et al., 2022b; Qiang et al., 2019; Ren et al., 2022; Ye, Yuwei et al., 2020; Ye, Y. et al., 2020). To fabricate CDs-based corrosion inhibitors, diverse one-step methods such as hydrothermal/solvothermal synthesis (Cen et al., 2019; Cui et al., 2017; Long et al., 2022b; Qiang et al., 2019; Ren et al., 2022; Ye, Yuwei et al., 2020; Zhao et al., 2022; Zheng et al., 2023), pyrolysis reaction (Ye, Y. et al., 2020), chemical oxidation method (He et al., 2022a) and Schiff base reaction (Long et al., 2022a) have been developed. Among them, by virtue of facile manipulation, easily available raw materials and no need of special equipment, hydrothermal/solvothermal approach has acted as the preferred way to synthesize

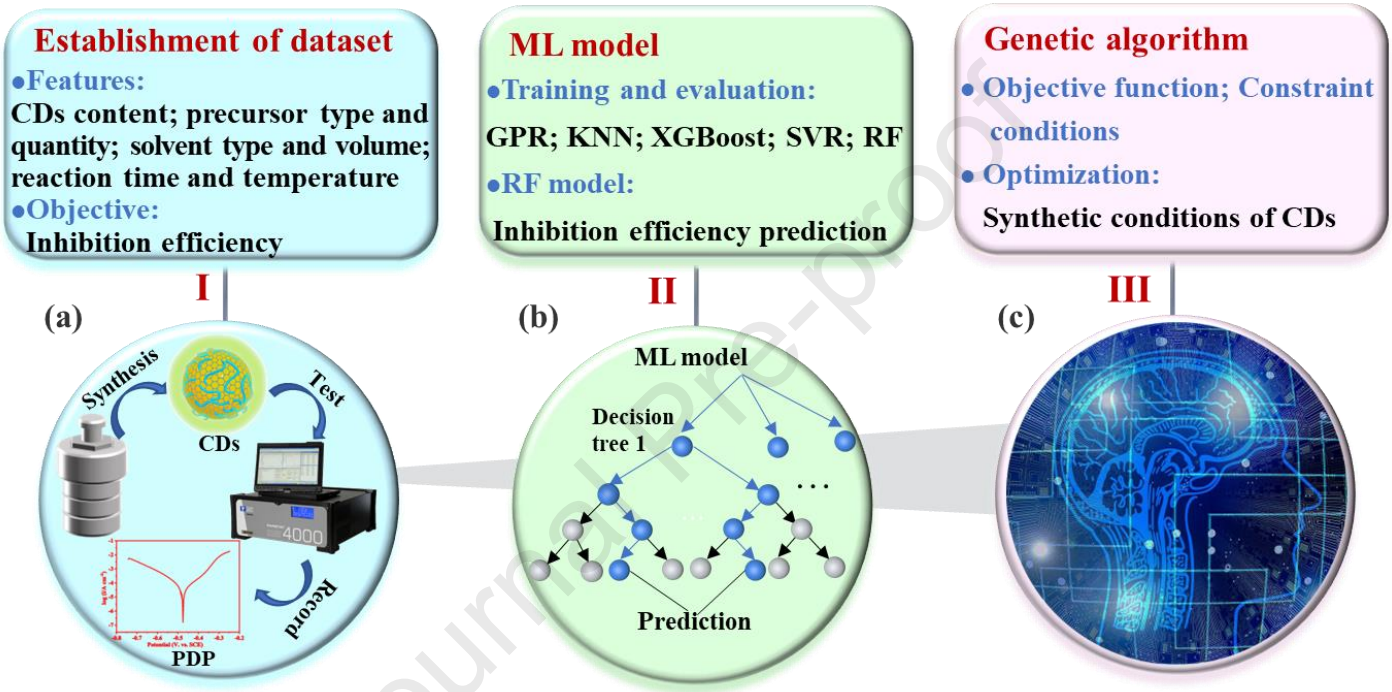
CDs (He et al., 2022b; Wang et al., 2022). Therefore, to efficiently obtain CDs with the high inhibition efficiency, identifying hydrothermal reaction parameters affecting the inhibition performance of CDs is urgently necessary.

At present, the trial-and-error approach is widely employed to prepare CDs-based corrosion inhibitors. Since this traditional method is driven by artificial experience, it suffers from inherent shortcomings including randomness, subjectivity and contingency. In detail, to obtain CDs with desirable inhibition efficiency, a large number of random synthetic experiments and screening processes are required, resulting in a low synthesis efficiency. Even for the most experienced researchers, they also cannot directly provide reasonable hydrothermal reaction parameters to generate CDs with the desired inhibition characteristic. For example, He's group (He et al., 2022a) used the green hydrothermal treatment to synthesize the biomass-derived CDs as an efficient corrosion inhibitor. And Guo's group (Zhu et al., 2022) adopted a hydrothermal method to obtain N-doped CDs, and demonstrated the outstanding inhibition behavior of these CDs for carbon steel. Although the reported studies above successfully prepare the different CDs-based inhibitors, the hydrothermal parameters are randomly selected. Thus, there is a high probability that the formed CDs do not possess the best inhibiting performance. If CDs-based corrosion inhibitors are only blindly prepared without predictable design and controlled synthesis, it will undoubtedly lead to invalid synthetic experiments, tremendous costs and potential environmental risk. As is well-known, the characteristics of the final CDs products are determined by precisely defined synthetic parameters (Wang et al., 2021). However, the hidden rule between hydrothermal synthesis conditions and inhibition behaviors of CDs is still unknown. This gives rise to failing to controllably prepare CDs-based corrosion inhibitors with the predictable inhibition efficiency. As a result, it is highly urgent to develop a feasible strategy to comprehensively reveal the correlation between hydrothermal reaction parameters and inhibition performance of CDs, thereby achieving predictable design

and controlled synthesis of CDs-based corrosion inhibitors.

As a subtopic of artificial intelligence technology, machine learning (ML) has made enormous progress in recent years, which can efficiently process large amounts of experimental and computationally simulated data to fully mine the hidden information in the data. This superiority enables ML to be widely applied in various research fields including strength prediction (Li and Song, 2023), damage evaluation (Ai et al., 2022), structural health monitoring (Ai et al., 2023), chemical reactions monitoring (Meuwly, 2021), transportation network detection (Sun et al., 2023), imaging analysis (Flah et al., 2020; Hao et al., 2023) and materials performance optimization (Tran and Ulissi, 2018). Especially for the development of novel materials, ML with a strong adaptability shows high prediction accuracy to their properties through effectively learning from the past and even failed data, immensely expediting the exploration period and reducing the cost. More importantly, the feasibility and potential of introducing ML into the field of CDs research have been convincingly proved (Chen et al., 2022; Han et al., 2020; Hong et al., 2022; Luo et al., 2022; Wang et al., 2021). For example, in 2020, Wu's group developed an XGBoost ML model to reveal the relationship between various synthesis parameters and experimental outcomes, resulting in the successful creation of green-emission CDs with fluorescent quantum yields of up to 39.3% (Han et al., 2020). Subsequently, Qian's group also used an extreme gradient boosting (XGBoost) ML model to fabricate customized CDs with excellent optical properties (Hong et al., 2022). More recently, Liu's group created a Random Forest (RF) ML model to investigate the reaction parameters and photoluminescence properties of CDs in multiple dimensions, accurately predicting their photoluminescence properties (Chen et al., 2022). And Huang’s group revealed the feasibility of ML to assist the synthesis of red fluorescent CDs (Luo et al., 2022). Existing studies employing ML to predict the optical properties of CDs have made significant progress, but they are unable to achieve intelligent optimization preparation of CDs, resulting in the inability to controllably synthesize CDs with

specific properties (Chen et al., 2022; Han et al., 2020; Hong et al., 2022; Luo et al., 2022; Wang et al., 2021). More importantly, no research on the predictable design and controlled preparation of CDs-based corrosion inhibitors *via* ML has been reported until now. This research gap makes it impossible to achieve controlled synthesis of CDs-based corrosion inhibitors based on ML, seriously hindering the development of CDs-based corrosion inhibitors.



Scheme 1 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.

Taking this into account, the current work aims to provide a suitable ML model to predict inhibition efficiency and optimize the synthesis process of CDs-based corrosion inhibitors, resulting in their controlled synthesis. The demerits of the trial-and-error method including randomness, subjectivity and contingency are successfully surmounted. In detail, a RF ML regression model on hydrothermally synthesized CDs-based corrosion inhibitors is established, revealing the relationship between various synthesis parameters and the inhibition efficiency of CDs. Guided by the RF model, the inhibition efficiencies of CDs-based corrosion

inhibitors are precisely predicted, and their synthesis route is intelligently optimized based on the genetic algorithm (GA) (**Scheme 1**). In the Section of Results and discussion, firstly, the processes for predictable design and controlled preparation of CDs-based corrosion inhibitors by ML are depicted; Secondly, extraction of the dataset for ML is presented; Subsequently, construction of ML model for inhibition efficiency prediction is shown; And then, intelligent optimization of the synthesis route for CDs-based corrosion inhibitors is illustrated; Finally, characterization of CDs-based corrosion inhibitors synthesized with the assistance of the GA is performed. This study demonstrates that ML as an effective strategy can establish the relationship between synthesis parameters and inhibition efficiencies of CDs, and guide the optimization of synthesis conditions for CDs-based inhibitors. All the observations above would assist researchers in the controlled synthesis of CDs-based corrosion inhibitors with predictable inhibition efficiency.

2. Experimental section

2.1. Materials and Reagents

Concentrated hydrochloric acid (HCl; AR grade, 36%~38%) was purchased from Sinopharm Chemical Reagent Co., Ltd. (Shanghai, China). Ammonium citrate (AC; AR grade), *o*-phenylenediamine (*o*-PD; AR grade) and citric acid (CA) were collected from Shanghai Macklin Biochemical Co., Ltd. (Shanghai, China). All the chemicals were used as received without further purification. Deionized (DI) water was adopted throughout the experiment. Q235 carbon steel specimens with the size of 15 mm × 15 mm × 3 mm were obtained from the local supplier. They were polished using several grades of emery papers (400, 600, 800, 1000, 1200, 1500 and 2000 grid) sequentially, then washed by DI water and degreased through ethanol, finally dried under cold air flow for electrochemical experiments.

2.2. Synthesis of CDs-based corrosion inhibitors

The precursor (AC, *o*-PD or CA) and DI water with a certain mass ratio were sealed into a Teflon-lined

stainless autoclave, and then placed into a drying oven at some temperature for a certain time. After cooling to room temperature, the mixture was centrifuged at 8000 rpm for 15 min, and then filtered *via* a 0.1 μm membrane to remove large particles. Finally, to obtain solid-state CDs, rotary evaporation and vacuum freeze-drying were used to treat mixture above. The collected CDs could be directly used as the corrosion inhibitors. Hence, in this work, CDs-based corrosion inhibitors refer to CDs.

2.3. Characterization of CDs

Transmission electron microscopy (TEM, Hitach HT7700, Japan) were used to observe CDs morphology. Raman spectra were confirmed using an alpha300 R Raman spectrometer (WITec, Germany) with an excitation wavelength of 488 nm. Fourier transform infrared spectroscopy (FTIR, Thermo Scientific Nicolet-6700, USA) and X-ray photoelectron spectrometer (Thermo Scientific K-Alpha+, USA) were introduced to confirm the functional groups information of CDs. The Ultraviolet–visible (UV–vis) spectra of CDs were gathered by employing a Lambda 750 UV−Vis−NIR spectrophotometer (Perkin Elmer, USA) with a 1.0 cm optical path.

2.4. Electrochemical measurement

The test solutions for confirming inhibition efficiency of CDs were obtained as follows. A certain quality of CDs was directly added to 1 M HCl solution, and a homogeneous solution as the test solution was formed under the glass rod stirring. The electrochemical measurement containing electrochemical impedance spectra (EIS) and potentiodynamic polarization (PDP) curves was performed on a PARSTAT 4000+ electrochemical station with a classical three-electrode system, in which Q235 carbon steel sample, a platinum plate and a saturated calomel electrode (SCE) were used as the working electrode (WE), counter electrode (CE) and reference electrode (RE) respectively. WE with the exposure area of 1 cm^2 was placed into a self-made container. Prior to each measurement, open circuit potential (OCP) must attain a stable state. The testing

parameters for EIS were from 10^5 to 10^{-2} Hz with 10 mV amplitude sinusoidal voltage. PDP curves were measured by scanning the potential range within ±250 mV versus OCP with a scanning rate of 1 mV s^{-1}. All the electrochemical experiments were carried out three times to guarantee the reproducibility.

2.5. Random Forest (RF) model

The RF model is an ensemble learning algorithm introduced by Breiman and Cutler in 2001 (Breiman, 2001), combining the Bagging algorithm and decision tree algorithm. It offers a powerful and novel method for establishing a nonlinear mapping between input variables and their responses (Breiman, 2001). The RF model, able to categorize and predict data, works by constructing numerous decision trees from distinct subsets of the dataset and features. On account of its high prediction accuracy, robustness against outliers and noise, fast computation speed, and the ability to mitigate overfitting, the RF model is particularly suitable for managing huge data. The underlying principle of RF involves employing the bootstrap technique to generate random samples (known as bagging) from the original dataset, constructing individual decision trees, and selecting random feature subsets for splitting at each tree node. These decision trees are then combined to yield the final prediction through majority voting. The construction of each tree can be regarded as a partitioning of the data space, with the leaf nodes representing the partitions. These partitions are super-rectangular units that collectively cover the entire data space. The key steps include sample bootstrap sampling, random feature subspace selection, and majority voting (Breiman, 1996).

2.6. Genetic algorithm (GA)

GA, an evolutionary algorithm extensively used in computational mathematics (Whitley, 1994), delivers excellent answers to optimization issues. Inspired by evolutionary biology concepts such as heredity, mutation, natural selection, and hybridization, GA is well-suited for tackling complex optimization challenges. In an optimization problem, candidate solutions, referred to as individuals, are depicted as chromosomes. The

population evolves over successive generations towards improved solutions (Lim et al., 2004). The evolutionary process begins with an initial population consisting of randomly generated individuals and proceeds iteratively. Each generation entails evaluating the fitness of the entire population and randomly selecting a subset of individuals based on their fitness. These selected individuals are subjected to natural selection and mutation, resulting in the generation of a new population. This new population subsequently becomes the current population for the next iteration of the algorithm.

3. Results and discussion

3.1. Processes for predictable design and controlled preparation of CDs-based corrosion inhibitors by ML

As shown in **Scheme 1**, the processes for controlled preparation of CDs-based corrosion inhibitors with the predictable inhibition efficiency guided by ML are divided into the following three steps. Firstly, 102 data points on CDs synthesis and inhibition efficiency are collected from a large number of reported studies and our own experiments to create the dataset. And the features include CDs concentration in HCl, precursor type and quantity, solvent type and volume as well as reaction time and temperature, while the objective is the inhibition efficiency of each CDs-based corrosion inhibitor calculated by PDP serves. Secondly, after training and evaluation of various models, the RF regression model with the best performance is chosen as the optimal model, which can accurately conjecture the inhibition efficiency of the specific CDs-based corrosion inhibitor. Thirdly, the synthesis route of CDs-based corrosion inhibitors is intelligently optimized by GA, based on the RF model and with the highest corrosion inhibition efficiency as the optimization objective. These procedures will be thoroughly explained in following sections.

3.2. Extraction of the dataset for ML

N doping endows CDs with more excellent inhibiting effects owing to its strong adsorption properties (Cen et al., 2019; Guo et al., 2022; Long et al., 2022a; Zhu et al., 2022). Hence, the employment of N-

containing precursors conduces to improving the inhibition efficiency of CDs. The obvious regularity above is easily observed from previous studies, but it is quite difficult to explicate clearly the association between synthesis conditions of CDs and their corresponding inhibition properties. In this case, it is necessary to use ML to acquire hidden information behind the synthesis parameters, assisting researchers quickly choose the suitable reaction conditions. First of all, the dataset is created by combining 102 data points on CDs synthesis and inhibition efficiency from multitudes of reported studies and our own experiments. In detail, according to previous experience and existing research, CDs concentration in HCl, precursor type and quantity, solvent type and volume, reaction time and temperature possibly governing the inhibition performance of CDs are chosen as typical parameters. Since the precursor type and solvent type fail to directly input into the ML model, the precursor type is converted into two primary characteristics including N and C atomic content, while the solvent type is replaced by code figures (Number 1, 2, 3 and 4 represent ethanol, NaOH solution, distilled water and DI water, respectively). Naturally, these eight descriptors containing CDs concentration in HCl, N atomic content, C atomic content, precursor quantity, solvent type and volume, reaction time and temperature are regarded as features (**Table 1**). The inhibition efficiencies of CDs calculated by PDP act as the objective. It should be pointed out that the inhibition efficiencies of CDs are acquired only from the references utilizing CDs as corrosion inhibitors for Q235 carbon steel in 1 M HCl or 0.5 M H_2SO_4 at room temperature. In addition to extracting data of CDs from the published literature, 10 sets of data containing CDs synthetic conditions and inhibition efficiencies are collected by our experiments to perfect the dataset for ML. This contributes to ensuring the reliability of ML models in guiding the synthesis of CDs-based corrosion inhibitors. The inhibiting efficiencies of the as-prepared CDs at different concentrations are calculated by their PDP curves. The corresponding synthetic parameters of CDs and their inhibition efficiencies calculated by PDP are shown in **Table S1**. Subsequently, the pairwise correlation between different features is examined before ML model

training. As presented in **Fig. 1a**, all the pairs of features illustrate low linear correlations, testifying the validity of feature selection.

Table 1 Descriptions of the input features for RF.

Features	Descriptions
CDs concentration in HCl	The concentration of CDs-based corrosion inhibitors in 1 M HCl
N atomic content	The atomic mass of N multiplies by the number of N in the precursor
C atomic content	The atomic mass of C multiplies by the number of C in the precursor
Precursor quantity	The total quantity of precursor used to prepare CDs
Solvent type	The type of solvent employed to synthesize CDs
Solvent volume	The total volume of solvent used to prepare CDs
Reaction time	Hydrothermal time for preparing CDs by hydrothermal method
Reaction temperature	Hydrothermal temperature for preparing CDs by hydrothermal method

3.3. Construction of ML model for inhibition efficiency prediction

The optimal ML model selection process is displayed in **Fig. S1**. After the establishment of the dataset, a series of different regression models containing gaussian process regression (GPR), K-nearest neighbor (KNN), XGBoost, support vector regression (SVR) and RF are examined by the established dataset. And 10-fold cross-validation is chosen as the evaluation method for these models, owing to its accurate and objective evaluation of model performance. When the requirements of the evaluation indicators containing root-mean-square error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) are met, the optimal model is screened to conjecture the inhibiting properties of CDs-based corrosion inhibitors, resulting in more efficient whole learning process. More specifically, the dataset is randomly shuffled, and then grouped at a ratio of 9:1 acting as the training set and validation set respectively. The former is applied to train the ML model; While the latter is employed to evaluate the performance of the thoroughly optimized ML model, thereby avoiding its overfitting. Subsequently, RMSE, MAE and R^2 are introduced to access the performance of different models. According to the conditions of the lowest RMSE and MAE as well as the highest R^2 in

the training set and validation set, the RF model outperforms the other models in predicting the inhibiting performance of CDs-based corrosion inhibitors, as depicted in **Fig. 1b** and **1c**. As is well known, RF model possesses high prediction accuracy, robustness against outliers and noise, fast computation speed and the ability to mitigate overfitting. It should be mentioned that by grid search, hyperparameter tuning is carried out to perfect these models. As a result, the RF model with an excellent generalization ability is finally chosen for further analysis and application.

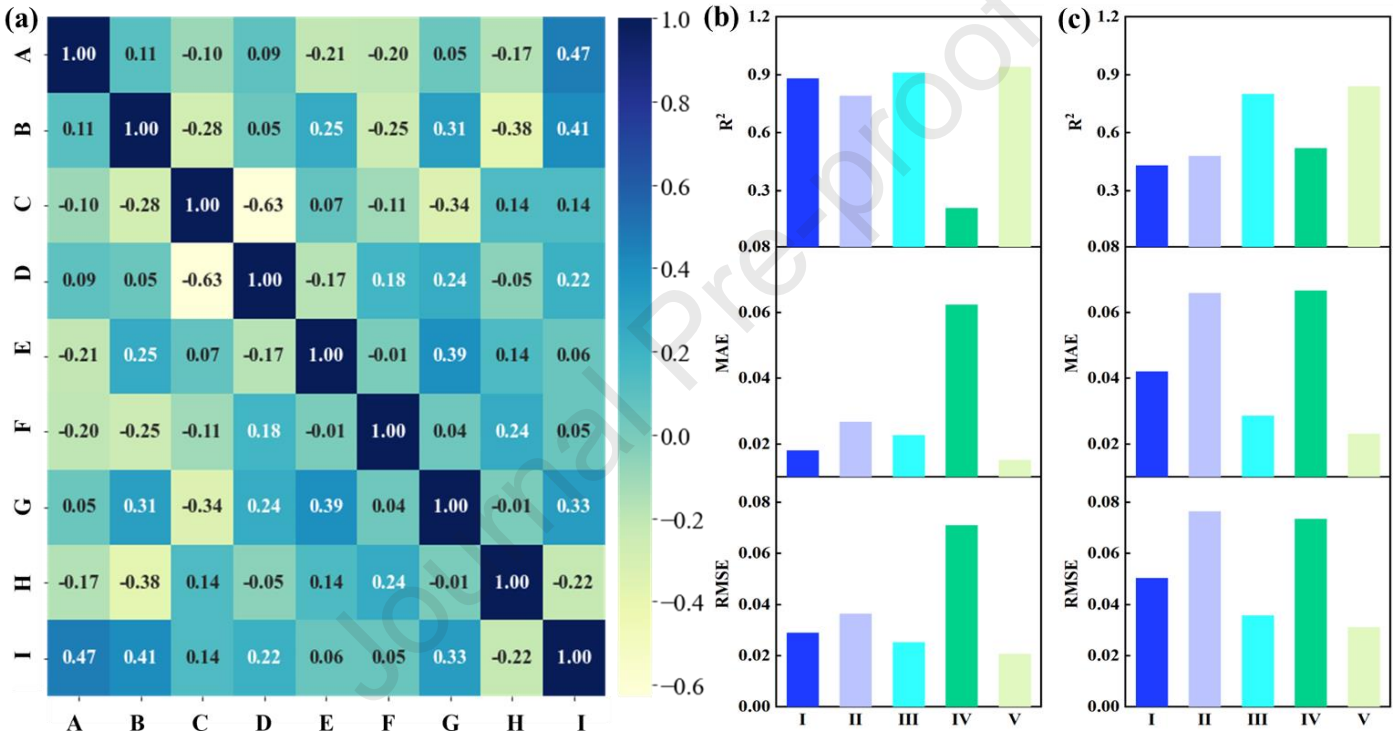


Fig. 1 (a) Heat map of Pearson's correlation coefficient matrix among the selected features of CDs (from A to I: CDs concentration in HCl, N atomic content, reaction time, solvent type, solvent volume, precursor quantity, C atomic content, reaction temperature and inhibition efficiency); Performance of different ML models evaluated by RMSE, MAE and R^2 in the (b) training set and (c) validation set (from I to VI: GPR, KNN, XGBoost, SVR and RF).

The performance of the RF model depends on the number of decision trees it includes. To determine the optimal number of decision trees, the RF model is tested with the number of trees ranging from 200 to 1000.

The results reveal that the Out of Bag (OOB) error consistently decreases until the number of trees reaches 460. Beyond this point, the classification error does not change significantly, even as the number of trees increases up to 1000. Therefore, the RF structure developed in this study consists of 460 decision trees. To display the effectiveness and generalization performance of the trained RF model, it is necessary to prove its performance in the predicting inhibiting behaviors of CDs-based corrosion inhibitors. As illustrated in **Fig. 2a** and **2c**, regardless of training or validation set, the majority of experimental values fall on the line y = x, indicating that the predicted values are in good agreement with the true values. The error rate (that is relative error between experimental and predicted values) leads to the same conclusion as above, since the error rates for the training and validation sets are almost less than 10% (**Fig. 2b** and **2d**), and synchronously, the prediction error (namely, the absolute error between the predicted value and experimental value) for most of data is within ±0.02 (**Fig. S2**). The above findings verify that the established RF model possesses an eminent feature extraction ability, converting the data into better feature expression and learning the distribution regularity of the data to accurately predict the inhibition performance of CDs-based corrosion inhibitors without overfitting. Consequently, the trained RF model with an outstanding generalization performance combines multiple data processing algorithms, which is suitable to effectively enhance the synthesis efficiency of CDs-based corrosion inhibitors. In detail, the designed synthesis parameters are simply input into the RF model prior to the implementation of the synthetic experiment of CDs-based corrosion inhibitors. The prediction result will then show whether the synthesis conditions are capable of producing CDs-based corrosion inhibitors with the desired inhibition performance. This will greatly improve the synthesis efficiency of CDs-based corrosion inhibitors by avoiding invalid experiments. In addition, it should be noted that with the expansion of the dataset

in the future, the generalization performance of the RF model will be further strengthened.

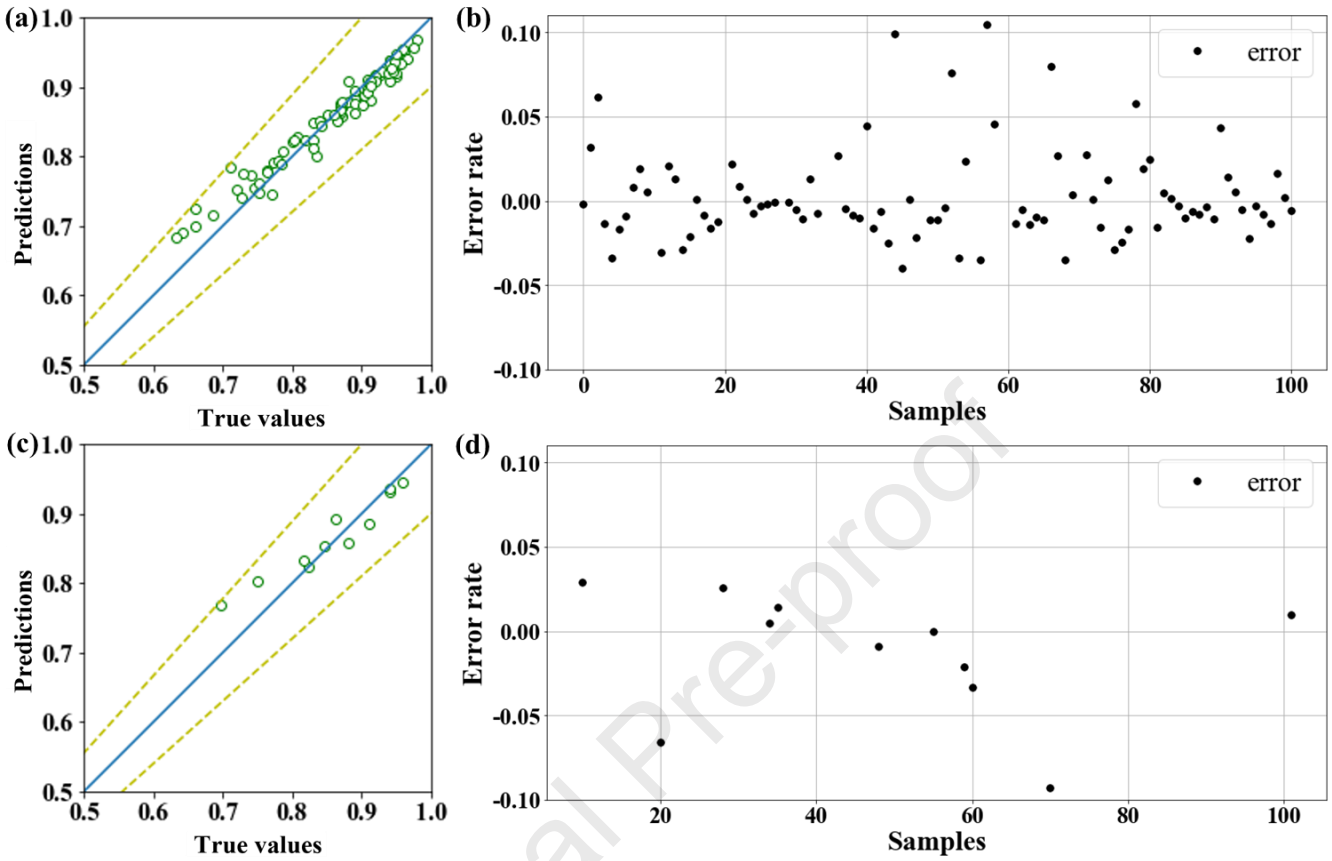


Fig. 2 The prediction effect for inhibiting behaviors of CDs-based corrosion inhibitors by the trained RF model: scatter plot and error plot of predicted and experimental values for (a-b) the training dataset and (c-d) test dataset.

To comprehensively reveal the correlation between hydrothermal reaction parameters and inhibition performance of CDs-based corrosion inhibitors, the importance of each feature is extracted from the trained RF model. The importance of a feature indicates how much it contributes to the prediction. The higher the importance of a feature, the greater the influence on the experimental outcome, that is, the value of inhibition efficiency. As indicated in **Fig. 3**, CDs concentration in HCl plays the most important role determining the inhibition behaviors of synthesized CDs, followed by N atomic content and reaction time. This result well coincides with the previous research (Cui et al., 2017; He et al., 2022a; Long et al., 2022a; Long et al., 2022b;

Qiang et al., 2019; Ren et al., 2022; Ye, Yuwei et al., 2020; Ye, Y. et al., 2020). Normally, regardless of the type of CDs, the higher the CDs content, the better the corrosion inhibition performance. This is possibly attributed to the fact that CDs with a higher concentration tend to form a more effective film on the surface of metals. Hence, CDs concentration in HCl is the most important feature to affect the value of inhibition efficiency. Moreover, from the inhibition mechanism aspect to interpret, N-containing groups give rise to the formation of coordination bonds between lone pair electrons of N atom and the unoccupied 3d orbital of Fe atoms, rendering CDs easily adsorb on steel substrate to resist corrosion (Cen et al., 2019; Guo et al., 2022; Long et al., 2022a; Zhu et al., 2022). Thus, the N atomic content in the precursor dominates the inhibiting performance of CDs through influencing the number of N-containing groups in final CDs. Regarding reaction time, it primarily affects the dehydration and condensation degrees of CDs in the reaction process, thereby mainly determine the structures of the final CDs (Ehrat et al., 2017; Shuang et al., 2020). Consequently, CDs with various structures have the different inhibition efficiencies. The above observations demonstrate that ranking importance offers an effective route to explore the significance of different reaction parameters in the synthesis of CDs-based corrosion inhibitors, assisting researchers in selectively synthesizing CDs-based corrosion inhibitors according to the influence degree of parameters. In a word, the trained RF model not only presents a remarkable prediction performance for the inhibiting efficiency of CDs-based corrosion inhibitors, but also offers a valuable algorithm for evaluating the significance of diverse features in the synthesis of CDs-based corrosion inhibitors, thus revealing the relationship between hydrothermal reaction parameters and their inhibition performance.

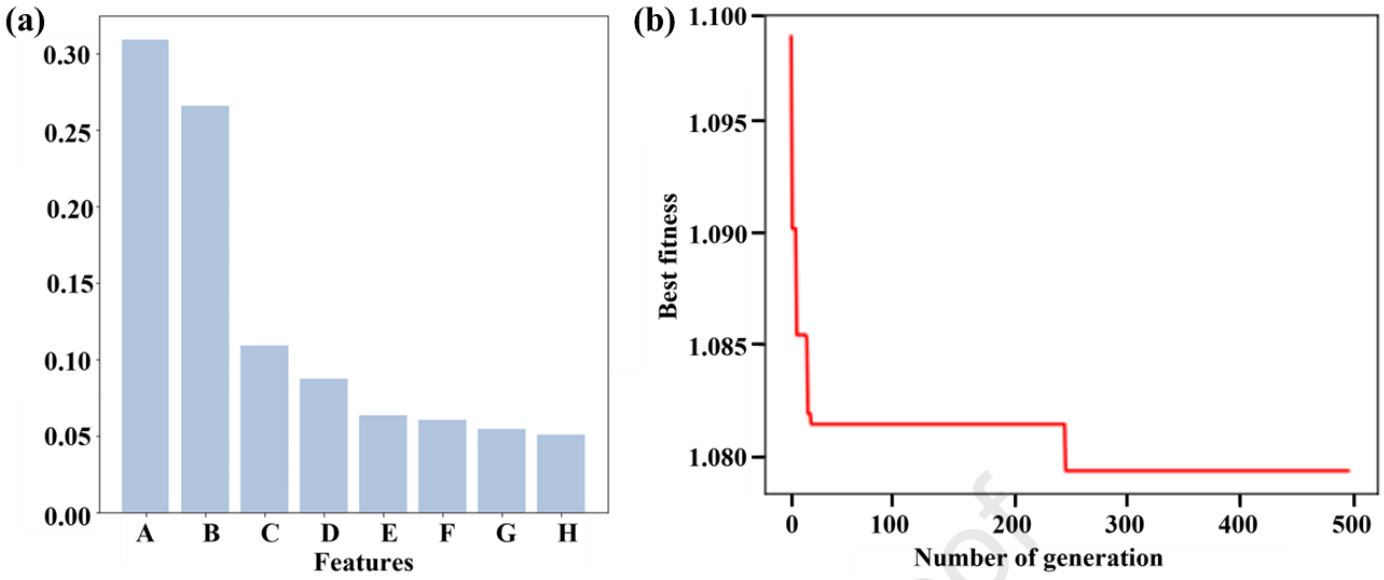


Fig. 3 (a) The feature importance extracted from the RF model (from A to H: CDs concentration in HCl, N atomic content, reaction time, solvent type, solvent volume, precursor quantity, C atomic content and reaction temperature); (b) the iterative curve for the optimization of inhibiting efficiency.

3.4. Intelligent optimization of the synthesis route for CDs-based corrosion inhibitors

GA is generally known as a way of searching for optimal solutions by simulating natural evolutionary processes, well-suited for addressing complex optimization issues. To dismiss the conventional trial-and-error exploration, the GA is introduced on the basis of the RF model to achieve intelligent optimization of the synthesis route for CDs-based corrosion inhibitors with the highest inhibition efficiency as the optimization objective. The optimization schematic with the highest inhibition efficiency as the objective is shown in **Fig. S3**. First of all, the RF model as the prediction model is established, whose creation process is described in Section 3.3. Then the developed RF model is used to calculate the inhibition efficiency, and the objective function is set as maximized inhibition efficiency. At the same time, the synthesis of CDs is subjected to various constraints, thus some constraint conditions are essential to be defined. With respect to the precursor, *o*-PD containing both carbon and nitrogen elements is frequently selected as the precursor to synthesize N-doped CDs. More importantly, *o*-PD derived N-doped CDs present a favorable inhibiting behavior for the steel in the corrosive acid solution. However, in the existing literature, the hydrothermal preparation conditions

of *o*-PD derived N-doped CDs still suffer from long preparation time, unfriendly solvents and requirement of additional precursors. Thus, it is highly urgent to optimize the preparation conditions of *o*-PD derived N-doped CDs without reducing their inhibition properties. Inspired by this, the *o*-PD is chosen as the precursor. If the preparation conditions of *o*-PD derived N-doped CDs is able to intelligently optimized by GA based on the RF model, it not only contributes to the efficient and low-cost use of *o*-PD, but also validates the application potential of ML in CDs-based corrosion inhibitors. Furthermore, in light of the previous reports and the established dataset in Section 3.1, the range constraint of each feature is displayed in **Table 2**. Finally, to obtain the highest inhibiting efficiency, GA optimization is performed, including the calculation process of selection, crossover and variation. And with the repetitions number condition, the optimization process is terminated until the final condition of the algorithm is satisfied (namely, convergence of the iteration curve). The typical iterative curve during the optimization process is described in **Fig. 3b**, presenting that the best fitness converges after around 240 iterations. After that, the optimized synthetic conditions are figured out. Meanwhile, taking into account the reduction of reaction time and the use of environmentally friendly solutions, the most appropriate optimized synthesis conditions are selected. That is, the amount of *o*-PD is 1.85 g; the solvent type is DI water; the solvent volume is 43.3 mL; the reaction time is 2.28 h, and the reaction temperature is 188°C.

Table 2 The requirements for the lower limit and the upper limit of each feature (Since the precursor type is *o*-PD, N and C atomic contents are confirmed).

Feature	Lower limit	Upper limit
CDs concentration in HCl (mg/L)	1	200
Solvent type	1	4
Solvent volume (mL)	0	100
Precursor quality (g)	0.6	12.7
Reaction time (h)	0.5	10
Reaction temperature (°C)	150	200

3.5. Characterization of CDs-based corrosion inhibitors synthesized with the assistance of the GA

Prior to confirming the inhibiting behavior of the product synthesized with the assistance of the GA, its morphological, structural and optical characterization is performed to testify the successful synthesis of CDs-based corrosion inhibitor (i.e., the as-prepared CDs). The morphology of the product is identified by TEM. As displayed in **Fig. 4a**, most of particles with a quasi-spherical shape share a favorable dispersion, and possess a size distribution of 1.41–2.73 nm in range and *ca*. 2.05 nm on average by randomly counting more than 100 particles. FTIR spectra and XPS are preformed to confirm the exhaustive information referring to surface chemical states and the elemental composition of the obtained product. For FTIR spectra (**Fig. 4b**), the broad absorption peaks at 3200−3400, 1630, 1459 and 1054 cm^{-1} are assigned to –OH/–NH, C=O, C–N and C–O respectively (E et al., 2021; Fu et al., 2023; Xia et al., 2018), suggesting the product possess numerous O- and N-containing functional groups, thereby conferring outstanding water solubility and excellent dispersity in aqueous systems on the product. The similar conclusion can be drawn from XPS. The full range XPS (**Fig. S4**) of CDs involves three marked peaks, assigned to C1s, N1s and O1s. High-resolution XPS C1s, O1s and N1s spectra (**Fig. 4c-e**) demonstrate the presence of C=C/C–C, C–O, C=O, C–N and $N–(C)_3$ (E et al., 2021; E et al., 2018; Li et al., 2019). The optical properties of CDs are measured by UV–vis spectra. As displayed in **Fig. 4f**, the obvious absorption peaks at 232 and 288 nm are ascribed to the π–π* transition of aromatic sp^2 domains (Xu et al., 2022), and the weak absorption peak at 421 nm is attributed to n−π* transition of O- and N-containing functional groups (Zhang et al., 2018). This implies that the particles in the product are composed of the carbon core as well as O- and N-containing surface groups, giving rise to multiple electronic transition states. In addition, the product aqueous solution with the concentration of 1 mg/mL shows yellow color exposed to visible light and emit a bright yellow fluorescence excited by a 365 nm UV light. According to the definition of CDs, which is that CDs normally are small carbon nanoparticles with the size less than 10

nm and the fluorescence as an instinct nature, and a single CD typically consists of a carbon core and oxygen/nitrogen-containing surface groups (Ai et al., 2021; He et al., 2021; He et al., 2022b), the above product is doubtlessly CDs (directly serving as the corrosion inhibitor). Therefore, the CDs are successfully prepared with the assistance of the GA.

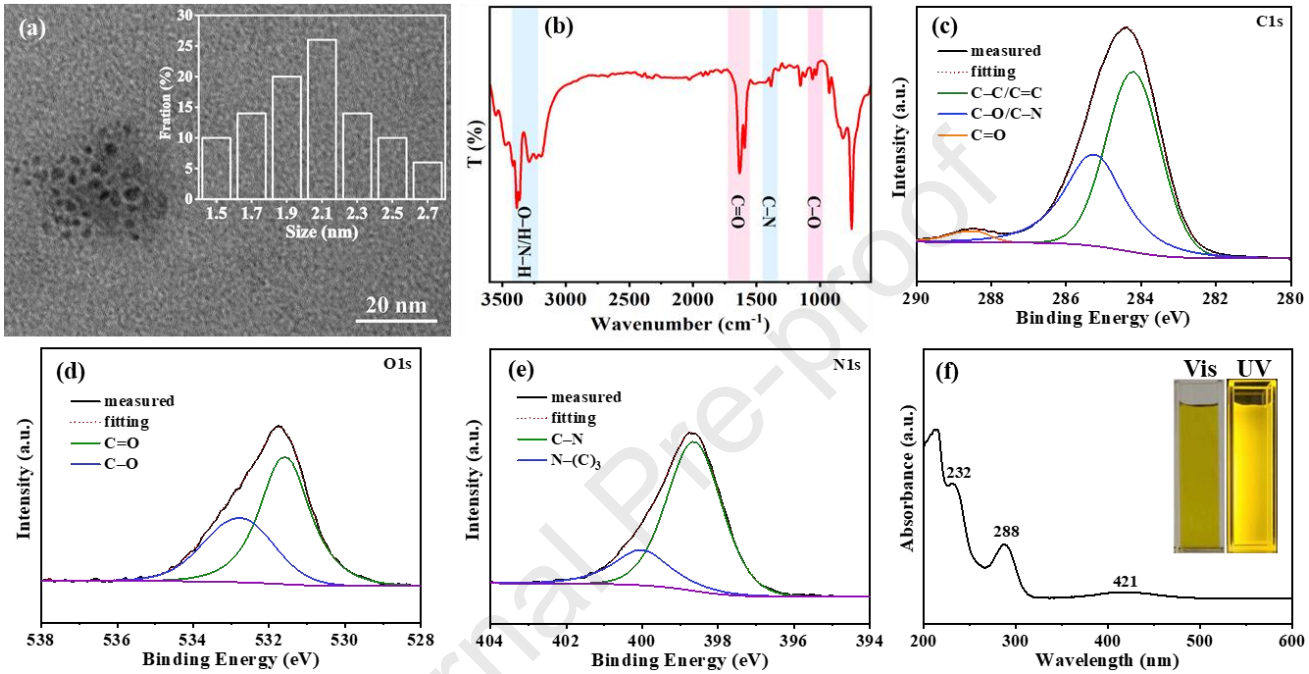


Fig. 4 Morphological and structural characterization of CDs: (a) TEM image (with the corresponding size distribution, inset); (b) FTIR spectra; (c-e) high-resolution XPS C1s, O1s and N1s spectra; (f) UV-vis spectra (with photographs taken under daylight and a 365 nm UV light, inset).

To validate the successful controlled fabrication of CDs-based corrosion inhibitors with predictable inhibition efficiency, the measured inhibition efficiency of as-prepared CDs is compared to their predicted inhibition efficiency originating from by the GA. PDP test is executed to find out the measured inhibition efficiency (IE) of CDs for Q235 carbon steel in 1 M HCl at room temperature (**Fig. 5**). The corrosion potential (E_{corr}), corrosion current density (i_{corr}), cathode slope (β_c), anode slope (β_a) and IE of blank group and CDs-containing group are summarized in **Table 3**. With respect to blank group, the CDs-containing group presents a positive shift in the E_{corr} value, suggesting the effective inhibition performance of the as-obtained CDs guided

by the GA. Clearly, both β_c and β_a values uncover a great decrease in the presence of CDs, and its β_a value reduces more sharply than the β_c value, implying that CDs demonstrate mixed-type inhibitive actions, and dramatically restrain the anodic dissolution reaction on the anode (He et al., 2022a). Correspondingly, the i_{corr} value of CDs-containing group decreases by two orders of magnitude compared to the blank group, leading to a high IE value of 92.3%. This measured IE (92.3%) exhibits excellent agreement with the predicted inhibition efficiency (94.6%) only within an error less than 3%. Similarly, these above findings can also be observed from the EIS measurement (**Fig. 5b**). By calculation, the inhibition efficiency value (η) in the presence of CDs boosts more than 96%, close to the predicted inhibition efficiency value (**Table 4**). And as depicted in Bode plots (**Fig. 5c**), the $|Z|_{0.01\ Hz}$ value of CDs groups increases tremendously in striking contrast to that of the blank group. At the same time, the phase angle in the presence of CDs dramatically enlarges, signifying the adsorption of CDs on the metal matrix (Ye et al., 2019). All the results above not only suggest that CDs synthesized with the assistance of the GA can efficiently suppress the corrosion of carbon steel in HCl solution, but also demonstrate that based on the GA, CDs-based corrosion inhibitors with a predictable inhibition efficiency are able to controllably synthesized. It must be emphasized that in comparison to the previous report using the same precursor (*o*-PD) to blindly prepare the CDs-based corrosion inhibitor (Cui et al., 2018), the reaction time in this work is enormously reduced by 81.0% (from 12.0 to 2.28 h) at the same reaction temperature, and the solvent is changed from ethanol to safer and cheaper DI water. More importantly, the two kinds of CDs share the similar inhibition efficiency. These comparisons further illustrate that based on the GA, ML can optimize the synthesis route of CDs with a desirable inhibiting performance, avoiding invalid synthetic experiments driven by trial-and-error method and tremendously enhancing the synthetic efficiency of CDs-based corrosion inhibitors.

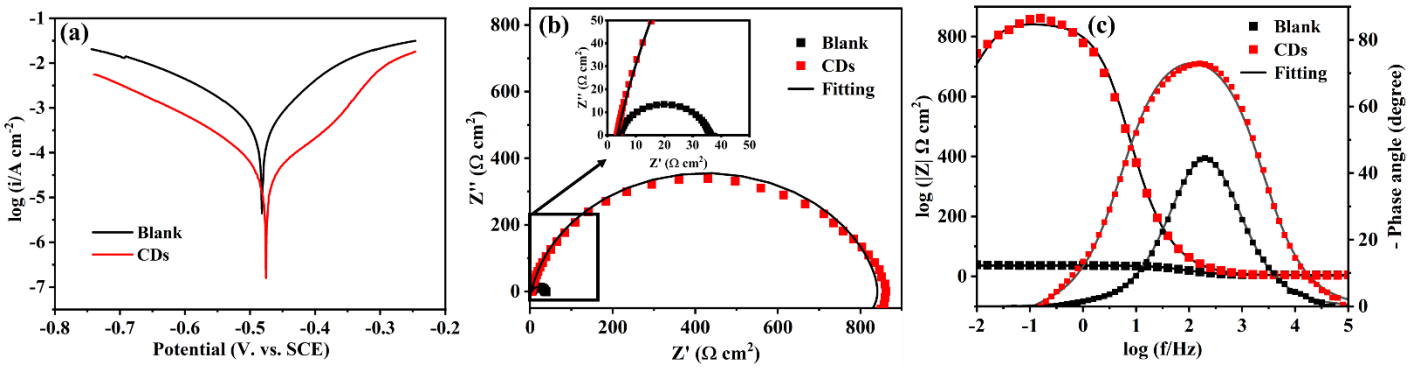


Fig. 5 Electrochemical characterization: (a) PDP curves (b) Nyquist and (c) Bode plots.

Table 3 The calculated electrochemical parameters from PDP curves.

Sample	E_{corr} (mV/SCE)	i_{corr} (μA/cm^2)	β_c (mV/dec)	β_a (mV/dec)	IE (%)
Blank	-482	534.2	130	97	
CDs	-475	41.0	98	94	92.3

Table 4 The fitting electrochemical parameters from EIS plots.

Sample	R_s (Ω cm^2)	R_{ct} (Ω cm^2)	CPE_{dl} (μF cm^{-2})	n	η (%)
Blank	4.4	32.0	158	0.890	
CDs	3.5	826.4	52	0.894	96.1

4. Conclusions

To cover the gap regarding the lack of controllable preparation for CDs-based corrosion inhibitors, the present study firstly exploits the RF ML model to successfully guide the controlled synthesis of CDs-based corrosion inhibitors with precisely predictable inhibiting efficiency, overcoming defects of the trial-and-error way. The main conclusions were summarized as follows.

(1) The dataset is established by extracting 102 synthesis data points from multitudes of reported studies and our experiments. Then the RF regression model with the optimal performance is created, which can accurately predict the inhibition efficiencies of CDs-based corrosion inhibitors with an error less than 10%.

(2) With the assistance of feature importance from the RF model, the decisive factors in the synthesis of CDs-based corrosion inhibitors can be screened out. The concentration of CDs in HCl has the greatest influence on the inhibitory behaviors of the synthesized CDs, followed by N atomic content and reaction time.

This contributes to better understanding the relationship between various hydrothermal synthesis parameters and the inhibition efficiency.

(3) Guided by the RF model, the synthesis route of CDs-based corrosion inhibitors is intelligently optimized by employing GA. The error between the measured IE (92.3%) and the predicted inhibition efficiency (94.6%) is only less than 3%. Therefore, controlled preparation for CDs-based corrosion inhibitors with a favorable and predictable inhibition efficiency is successfully achieved.

These findings provide an effective ML strategy to controllably synthesize CDs-based corrosion inhibitors with the predictable inhibition efficiency, filtering out unsatisfactory synthesis conditions and greatly improving the synthetic efficiency of CDs-based corrosion inhibitors. Consequently, it contributes to reducing resource waste and minimizing environmental implications, supporting the sustainability and cleaner production in the field of corrosion inhibitors. In addition, the proposed ML is expected to be employed for the design and synthesis of other novel materials. It should be emphasized that the availability of high-quality data poses a challenge to the performance and generalization capabilities of ML model. Hence, the availability of diverse and well-curated datasets encompassing a wide range of corrosion inhibitors and experimental conditions would conduce to the robustness and accuracy of inhibition efficiency predictions. In this work, the dataset is constructed by combining 102 data points from numerous published studies and our own experiments. Future efforts should be made to collect high-quality data to further refine and update the as-exploited ML model. This ongoing improvement will enhance the accuracy and applicability of our ML approach, and foster the development of more efficient and sustainable corrosion inhibition strategies.

Acknowledgments

This work is financially supported by the National Natural Science Foundation of China-Youth Science Fund (No. 52208273), the Dalian Science and Technology Talent Innovation Support Program (No. 2022RQ035),

the Exploration Project of the Natural Science Foundation of Zhejiang Province (No. LQ21E080003 and No. LTZ22D010001) and the National Natural Science Foundation of China (No. 51978435). Furthermore, the authors thank Xinfang Cui from Shiyanjia Lab (www.shiyanjia.com) for Raman and XPS measurement.

References

Ai, L., Soltangharaei, V., Ziehl, P., 2022. Developing a heterogeneous ensemble learning framework to evaluate Alkali-silica reaction damage in concrete using acoustic emission signals. Mech. Syst. Signal Pr. 172, 108981.

Ai, L., Yang, Y., Wang, B., Chang, J., Tang, Z., Yang, B., Lu, S., 2021. Insights into photoluminescence mechanisms of carbon dots: advances and perspectives. Sci. Bull. 66(8), 839-856.

Ai, L., Zhang, B., Ziehl, P., 2023. A transfer learning approach for acoustic emission zonal localization on steel plate-like structure using numerical simulation and unsupervised domain adaptation. Mech. Syst. Signal Pr. 192, 110216.

Astuti, P., Rafdinal, R.S., Yamamoto, D., Andriamisaharimanana, V., Hamada, H., 2022. Effective Use of Sacrificial Zinc Anode as a Suitable Repair Method for Severely Damaged RC Members Due to Chloride Attack. Civ. Eng. J. 8(7), 1535-1548.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123-140.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5-32.

Cen, H., Chen, Z., Guo, X., 2019. N, S co-doped carbon dots as effective corrosion inhibitor for carbon steel in CO2-saturated 3.5% NaCl solution. J. Taiwan Inst. Chem. E. 99, 224-238.

Chen, J., Luo, J.B., Hu, M.Y., Zhou, J., Huang, C.Z., Liu, H., 2022. Controlled Synthesis of Multicolor Carbon Dots Assisted by Machine Learning. Adv. Funct. Mater. 33(2), 2210095.

Cui, M., Ren, S., Xue, Q., Zhao, H., Wang, L., 2017. Carbon dots as new eco-friendly and effective corrosion inhibitor. J. Alloy. Compd. 726, 680-692.

Cui, M., Ren, S., Zhao, H., Wang, L., Xue, Q., 2018. Novel nitrogen doped carbon dots for corrosion inhibition of carbon steel in 1 M HCl solution. Appl. Surf. Sci. 443, 145-156.

E, S., He, C., Wang, J.H., Mao, Q., Chen, X., 2021. Tunable Organelle Imaging by Rational Design of Carbon

457 Dots and Utilization of Uptake Pathways. ACS Nano 15(9), 14465-14474.
458 E, S., Mao, Q.X., Yuan, X.L., Kong, X.L., Chen, X.W., Wang, J.H., 2018. Targeted imaging of the lysosome
459 and endoplasmic reticulum and their pH monitoring with surface regulated carbon dots. Nanoscale 10(26),
460 12788-12796.
461 Ehrat, F., Bhattacharyya, S., Schneider, J., Lof, A., Wyrwich, R., Rogach, A.L., Stolarczyk, J.K., Urban, A.S.,
462 Feldmann, J., 2017. Tracking the Source of Carbon Dot Photoluminescence: Aromatic Domains versus
463 Molecular Fluorophores. Nano Lett. 17(12), 7710-7716.
464 Flah, M., Suleiman, A.R., Nehdi, M.L., 2020. Classification and quantification of cracks in concrete structures
465 using deep learning image-based techniques. Cement Concrete Comp. 114, 103781.
466 Fu, R., Song, H., Liu, X., Zhang, Y., Xiao, G., Zou, B., Waterhouse, G.I.N., Lu, S., 2023. Disulfide
467 crosslinking-induced aggregation: towards solid-state fluorescent carbon dots with vastly different emission
468 colors. Chinese J. Chem. 41(9), 1007-1014.
469 Guo, L., Zhu, M., He, Z., Zhang, R., Kaya, S., Lin, Y., Saji, V.S., 2022. One-Pot Hydrothermal Synthesized
470 Nitrogen and Sulfur Codoped Carbon Dots for Acid Corrosion Inhibition of Q235 Steel. Langmuir 38(13),
471 3984-3992.
472 Han, Y., Tang, B., Wang, L., Bao, H., Lu, Y., Guan, C., Zhang, L., Le, M., Liu, Z., Wu, M., 2020. Machine-
473 Learning-Driven Synthesis of Carbon Dots with Enhanced Quantum Yields. ACS Nano 14(11), 14761-14768.
474 Hao, Z., Lu, C., Li, Z., 2023. Highly accurate and automatic semantic segmentation of multiple cracks in
475 engineered cementitious composites (ECC) under dual pre-modification deep-learning strategy. Cement
476 Concrete Res. 165, 107066.
477 He, C., E, S., Yan, H., Li, X., 2021. Structural engineering design of carbon dots for lubrication. Chinese Chem.
478 Lett. 32(9), 2693-2714.
479 He, C., Li, X.-Q., Feng, G.-L., Long, W.-J., 2022a. A universal strategy for green and in situ synthesis of
480 carbon dot-based pickling solution. Green Chem. 24(15), 5842-5855.
481 He, C., Xu, P., Zhang, X., Long, W., 2022b. The synthetic strategies, photoluminescence mechanisms and
482 promising applications of carbon dots: Current state and future perspective. Carbon 186, 91-127.
483 Hong, Q., Wang, X.-Y., Gao, Y.-T., Lv, J., Chen, B.-B., Li, D.-W., Qian, R.-C., 2022. Customized Carbon Dots
484 with Predictable Optical Properties Synthesized at Room Temperature Guided by Machine Learning. Chem.

Mater. 34(3), 998-1009.

Kobzar, Y.L., Fatyeyeva, K., 2021. Ionic liquids as green and sustainable steel corrosion inhibitors: Recent developments. Chem. Eng. J. 425, 131480.

Li, J., Gong, X., 2022. The Emerging Development of Multicolor Carbon Dots. Small 18(51), 2205099.

Li, J., Wu, Y., Gong, X., 2023. Evolution and fabrication of carbon dot-based room temperature phosphorescence materials. Chem. Sci. 14(14), 3705-3729.

Li, Q., Song, Z., 2023. Prediction of compressive strength of rice husk ash concrete based on stacking ensemble learning model. J. Clean. Prod. 382, 109401.

Li, W., Liu, Y., Wang, B., Song, H., Liu, Z., Lu, S., Yang, B., 2019. Kilogram-scale synthesis of carbon quantum dots for hydrogen evolution, sensing and bioimaging. Chinese Chem. Lett. 30(12), 2323-2327.

Lim, C.-H., Yoon, Y.-S., Kim, J.-H., 2004. Genetic algorithm in mix proportioning of high-performance concrete. Cement Concrete Res. 34(3), 409-420.

Long, W.-J., Li, X.-Q., Xu, P., Feng, G.-L., He, C., 2022a. Facile and scalable preparation of carbon dots with Schiff base structures toward an efficient corrosion inhibitor. Diam. Relat. Mater. 130, 109401.

Long, W.-J., Li, X.-Q., Yu, Y., He, C., 2022b. Green synthesis of biomass-derived carbon dots as an efficient corrosion inhibitor. J. Mol. Liq. 360, 119522.

Luo, J.B., Chen, J., Liu, H., Huang, C.Z., Zhou, J., 2022. High-efficiency synthesis of red carbon dots using machine learning. Chem. Commun. 58(64), 9014-9017.

Madlangbayan, M.S., Diola, C.N.B., Tapia, A.K.G., Peralta, M.M., Peralta, E.K., Almeda, R.A., Bayhon, M.A.L., Sundo, M.B., 2021. Corrosion Inhibition of Sodium Silicate with Nanosilica as Coating in Pre-Corroded Steel. Civ. Eng. J. 7(11), 1806-1816.

Meuwly, M., 2021. Machine Learning for Chemical Reactions. Chem. Rev. 121(16), 10218-10239.

Qiang, Y., Zhang, S., Zhao, H., Tan, B., Wang, L., 2019. Enhanced anticorrosion performance of copper by novel N-doped carbon dots. Corros. Sci. 161, 108193.

Ren, S., Cui, M., Chen, X., Mei, S., Qiang, Y., 2022. Comparative study on corrosion inhibition of N doped and N,S codoped carbon dots for carbon steel in strong acidic solution. J. Colloid Interface Sci. 628, 384-397.

Shuang, E., Mao, Q.X., Wang, J.H., Chen, X.W., 2020. Carbon dots with tunable dual emissions: from the mechanism to the specific imaging of endoplasmic reticulum polarity. Nanoscale 12(12), 6852-6860.

Sun, R., Luo, Q., Chen, Y., 2023. Online transportation network cyber-attack detection based on stationary sensor data. Transport. Res. C-Emer. 149, 104058.

Tan, B., Xiang, B., Zhang, S., Qiang, Y., Xu, L., Chen, S., He, J., 2021. Papaya leaves extract as a novel eco-friendly corrosion inhibitor for Cu in H2SO4 medium. J. Colloid Interface Sci. 582, 918-931.

Tran, K., Ulissi, Z.W., 2018. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. Nat. Catal. 1(9), 696-703.

Wang, B., Cai, H., Waterhouse, G.I.N., Qu, X., Yang, B., Lu, S., 2022. Carbon Dots in Bioimaging, Biosensing and Therapeutics: A Comprehensive Review. Small Sci. 2(6), 2200012.

Wang, X.Y., Chen, B.B., Zhang, J., Zhou, Z.R., Lv, J., Geng, X.P., Qian, R.C., 2021. Exploiting deep learning for predictable carbon dot design. Chem. Commun. 57(4), 532-535.

Whitley, D., 1994. A genetic algorithm tutorial. Stat. Comput. 4, 65-85.

Xia, J., Chen, S., Zou, G.Y., Yu, Y.L., Wang, J.H., 2018. Synthesis of highly stable red-emissive carbon polymer dots by modulated polymerization: from the mechanism to application in intracellular pH imaging. Nanoscale 10(47), 22484-22492.

Xu, Y., Wang, B., Zhang, M., Zhang, J., Li, Y., Jia, P., Zhang, H., Duan, L., Li, Y., Li, Y., Qu, X., Wang, S., Liu, D., Zhou, W., Zhao, H., Zhang, H., Chen, L., An, X., Lu, S., Zhang, S., 2022. Carbon Dots as a Potential Therapeutic Agent for the Treatment of Cancer-Related Anemia. Adv. Mater. 34(19), 2200905.

Ye, Y., Jiang, Z., Zou, Y., Chen, H., Guo, S., Yang, Q., Chen, L., 2020. Evaluation of the inhibition behavior of carbon dots on carbon steel in HCl and NaCl solutions. J. Mater. Sci. Technol. 43, 144-153.

Ye, Y., Yang, D., Chen, H., 2019. A green and effective corrosion inhibitor of functionalized carbon dots. J. Mater. Sci. Technol. 35(10), 2243-2253.

Ye, Y., Yang, D., Chen, H., Guo, S., Yang, Q., Chen, L., Zhao, H., Wang, L., 2020. A high-efficiency corrosion inhibitor of N-doped citric acid-based carbon dots for mild steel in hydrochloric acid environment. J. Hazard. Mater. 381, 121019.

Zhang, T., Zhao, F., Li, L., Qi, B., Zhu, D., Lu, J., Lu, C., 2018. Tricolor White-Light-Emitting Carbon Dots with Multiple-Cores@Shell Structure for WLED Application. ACS Appl. Mater. Interfaces 10(23), 19796-19805.

Zhao, H., Sun, T.-Y., Huang, L.-F., Wei, J., Qiu, S., 2022. A green strategy for nitrogen-doped polymer

nanodots with high oxygen and chloride corrosion resistance in extremely acidic condition. Chem. Eng. J. 437.

Zheng, S., Feng, L., Hu, Z., Li, J., Zhu, H., Ma, X., 2023. Study on the corrosion inhibition of biomass carbon quantum dot self- aggregation on Q235 steel in hydrochloric acid. Arab. J. Chem. 16(4), 104605.

Zhu, M., Guo, L., He, Z., Marzouki, R., Zhang, R., Berdimurodov, E., 2022. Insights into the newly synthesized N-doped carbon dots for Q235 steel corrosion retardation in acidizing media: A detailed multidimensional study. J. Colloid Interface Sci. 608, 2039-2049.

Highlights:

1. Machine learning is used to predict inhibition efficiency (IE) of carbon dots (CDs).

2. The correlation between synthesis parameters and IE of CDs is revealed.

3. Synthesis route of CDs-based corrosion inhibitors is optimized by genetic algorithm.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: