

Research Article

Open Access



Machine learning-assisted prediction, screen, and interpretation of porous carbon materials for high-performance supercapacitors

Hongwei Liu^{1,2,3,#}, Zhenming Cui^{1,2,3,#}, Zhennan Qiao^{1,2,3}, Xiaokang An^{1,2,3}, Yongzhen Wang^{1,2,3,*} 

¹College of Materials Science and Engineering, Taiyuan University of Technology, Taiyuan 030024, Shanxi, China.

²Shanxi Joint Laboratory of Coal-based Solid Waste Resource Utilization and Green Development, Taiyuan University of Technology, Taiyuan 030024, Shanxi, China.

³Innovation Research Center for Materials Genetic Engineering, Taiyuan University of Technology, Taiyuan 030024, Shanxi, China.

#Authors contributed equally.

*Correspondence to: Prof. Yongzhen Wang, College of Materials Science and Engineering, Taiyuan University of Technology, 79 Yingze West Main Street, Taiyuan 030024, Shanxi, China. E-mail: wangyongzhen@tyut.edu.cn

How to cite this article: Liu H, Cui Z, Qiao Z, An X, Wang Y. Machine learning-assisted prediction, screen, and interpretation of porous carbon materials for high-performance supercapacitors. *J Mater Inf* 2024;4:16. <https://dx.doi.org/10.20517/jmi.2024.29>

Received: 2 Aug 2024 **First Decision:** 14 Sep 2024 **Revised:** 1 Oct 2024 **Accepted:** 18 Oct 2024 **Published:** 24 Oct 2024

Academic Editor: Hao Li **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Porous carbon materials (PCMs) are preferred as electrode materials for supercapacitor energy storage applications due to their superior characteristics. However, the optimal performance of these electrodes requires trial and error experimental exploration due to the complexity of influencing factors. To address this limitation, we develop a machine learning (ML) combined experimental validation approach to predict, screen and interpret ideal PCMs for supercapacitors. Four ML models are used for predicting the specific capacitance (SC) properties of PCMs and the light gradient boosting machine (LGBM) model exhibits the best prediction performance with an R^2 value of 0.92. Through comprehensive interpretability analysis of ML, important variables influencing SC properties are identified and their impact range is determined. By analyzing the deviation of key values during experimental verification, accurate predictions of SC properties of PCMs are made, facilitating precise material screening. Additionally, the accuracy and applicability of the ML model predictions are evaluated. This research pioneered a key eigenvalue fall-point screening approach based on a combination of ML experiments for accurately predicting SC performance and screening of superior energy storage materials, providing a compelling strategy for advancing energy storage materials technology.



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



Keywords: Machine learning, capacitance predication, feature importance, Shapley additive explanation, interpretability, fall-point screening

INTRODUCTION

Supercapacitors play a pivotal role in advancing renewable energy technologies, thanks to their energy storage capacity^[1]. The energy storage capacity of supercapacitors is primarily influenced by the electrode materials employed^[2,3]. In the ongoing development of supercapacitors, various materials including porous carbon materials (PCMs), metal oxides, carbides, nitrides, and conductive polymers have been optimized to enhance energy storage capacity. Among these materials, PCMs remain the predominant choice as electrode active materials due to their power capability, long-term cycle stability, wide operating temperatures, and high coulomb efficiency^[4-6]. Previous studies have highlighted the importance of designing ideal materials with the necessary properties to improve the capacitive performance of supercapacitors^[7-9]. The key to the enhancement of capacitance performance is to improve the electric double-layer capacitance [affected or determined by specific surface area (SSA) and pore structure] and pseudo-capacitance induced by the redox reaction of heteroatoms [such as nitrogen (N), oxygen (O), or sulfur (S)]. Large surface area and appropriate pore structure of PCMs play a crucial role in providing significant interfacial areas, ionic mobility paths and charge transfer, thereby enabling the achievement of high specific capacitance (SC) and rapid charge/discharge rates^[10-13]. N or O, in particular, has been extensively researched due to its beneficial effects and relatively straightforward doping process. Specifically, N-doping into carbon matrix is the most promising approach to promoting the pseudo-capacitance performance and O-doping can improve the electrical conductivity and surface wettability, providing additional pseudo-capacitance. However, the complexity of synthesizing nitrogen or oxygen-doped PCMs leads to challenges in accurately identifying and controlling electrochemically active nitrogen or oxygen configurations. The current empirical trial and error approach involving the synthesis, characterization, testing, and analysis of various PCMs is time-consuming and resource-intensive. Optimizing or simplifying these processes is essential for efficiently screening PCMs and advancing the development of high-performance carbon electrode materials, necessitating the exploration of new process optimization methods.

Machine learning (ML) technology has become a valuable tool in accelerating material development and predicting performance^[14-16]. By analyzing data from previous research, ML methods can simulate and predict the relationship between input characteristics and output variables, identifying key features of porous carbon-based electrodes for supercapacitors^[17-19]. Various ML prediction models, such as artificial neural networks (ANN)^[20,21], random forest (RF)^[22], and eXtreme Gradient Boosting (XGBoost)^[23], have been used to predict the capacitive properties of PCMs. Researchers have successfully employed ML models to design high-performance supercapacitors by analyzing the relationship between capacitors and structural features of PCMs. XGBoost was found to deliver the best predictive performance among the ML models tested. Additionally, studies have shown that specific structural features, such as SSA (S_{mic}/SSA) and pore size (PS), significantly contribute to the capacitive performance of porous structures^[24]. Through the use of ML models such as k-nearest neighbors' regression (KNN), decision tree regression (DTR), Bayesian ridge regression (BRR), and ANN, Saad *et al.* have accurately predicted the capacitance of graphene-based supercapacitor electrodes based on physicochemical characteristics and electrochemical measurements^[25]. The ANN model, in particular, has demonstrated exceptional prediction accuracy, with the SHapley additive explanation (SHAP) framework revealing the significant influence of nitrogen and oxygen atomic percentages in doped graphene on the ANN model. Recent research has predominantly focused on predicting the capacitive performance of supercapacitors based on the structural characteristics of carbon electrode materials. However, there is a lack of comprehensive studies that consider the impact of doping species, porous carbon structure characteristics, and operational factors on capacitance performance.

Rahimi *et al.* elucidated the synergistic effect of N/O functional groups and microstructures of porous carbon on supercapacitor performance using a multilayer perceptron neural network (MLP-NN) model^[26]. The model demonstrated high accuracy in predicting porous carbon performance with minimal error. Through genetic algorithm optimization, the optimal SC of porous carbon was predicted to be 550 F·g⁻¹ (1 A·g⁻¹) in a three-electrode system. Sun *et al.* investigated the influence mechanism of 14 feature parameters on the energy storage characteristics of active biochar using various ML models^[27]. They identified key influencing characteristic parameters through functionality and partial correlation analysis, shedding light on the preparation of active biochar for supercapacitors. It can be seen that as an alternative, ML models can be utilized not only to analyze massive datasets gathered from experimental procedures, revealing quantitative relationships between key features of electrode materials and their impacts^[28], but also to interpret device performance based on theoretical modeling in non-equilibrium conditions. Furthermore, ML technology offers a valuable tool for the design of materials to get the utilization it deserves.

Therefore, based on the laboratory's previous experimental research on the preparation of nitrogen-oxygen co-doped PCMs derived from lignite (N, O co-doped PCMs) and the availability of literature data, this study selected the N, O co-doped PCMs as the representative carbon electrode material for supercapacitors. Seventeen key parameters related to physio-chemical and operational properties were gathered from recent literature and lab data on the preparation of porous carbon-based supercapacitor electrodes. Four ML models - linear regression (LR), RF, XGBoost, and light gradient boosting machine (LGBM) - were employed to investigate the impact of these parameters on the energy storage characteristics of N, O co-doped PCMs through computer simulations. The most effective prediction model was determined, and important features for accurately predicting supercapacitor characteristics were identified using characteristic analysis and partial dependence correlation analysis. A novel method for analyzing predicted and actual supercapacitor deviations of electrode materials was proposed. Experimental verification and model optimization were used to achieve performance prediction and accurate screening of N, O co-doped PCMs, along with the falling areas analysis method for key characteristics. This study presents an interpretable model for understanding the relationship between key characteristics and capacitive properties in N, O co-doped PCMs, providing a means to predict and screen these materials for high-performance supercapacitors.

MATERIALS AND METHODS

Dataset collection, processing, and feature selection

To predict energy storage characteristics of N, O co-doped PCMs, more than 400 raw datasets were collected based on N, O co-doped PCM samples. The partial database was based on our previous work. The required data were extracted from the experimental results of the studies carried out on N, O co-doped PCMs utilized in the electrode of supercapacitors. Moreover, the SC data at different current densities (0.5 and 1.0 A·g⁻¹) were abstracted from the published SCI journal papers, which were added to the original database. The data-gathering information on N, O co-doped PCM samples includes microstructure features, N or O content, N/O functional group intensity, and operational parameters. A total of 17 input features were gathered to predict the dataset of energy storage characteristics of N, O co-doped PCMs: the oxygen content (O%), nitrogen content (N%), as well as various functional groups containing N or O, including N-6, N-5, N-Q, the ratio of the carbonyl-O group (C=O, O-I), the ratio of hydroxyl-O/ether-O group (C-OH/C-O-C, O-II), the ratio of carboxyl-O group (-COOH, O-III), the SSA of N, O co-doped PCMs, micropore surface area (S_{mic}), micropore surface area proportion (S_{mic}/SSA), total pore volume (V_t), micropore volume (V_{mic}), pore size distribution (PSD), micropore volume proportion (V_{mic}/V_t), and the potential window (PW, V), and the current density (CD, A·g⁻¹) obtained from a SC test. The SC was selected as the output feature variable. All raw data from previous experimental and literature studies were performed in a 6 M KOH aqueous solution with a three-electrode system. Ultimately, 275 defect-free

datasets were screened with detailed values of inputs to ensure a more accurate evaluation and all modeling was conducted under the same electrolyte and system details.

As a first step of the analytical procedure, all considered variables were evaluated by the Pearson correlation coefficient (PCC), which was calculated by

$$PCC = \frac{\sum_{i=1}^n (x_i - \bar{x}) \sum_{i=1}^n |y_i - \bar{y}|}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Where x and y represent independent input and dependent output variables, and \bar{x} and \bar{y} are their mean values (x includes different features of N, O co-doped PCMs electrode and y for SC). The value of R can be between 1 and -1, where a positive or negative value means a positive or negative correlation, and 0 indicates a nonlinear correlation. The PCC was utilized to measure the linear dependence between all variables and applied to check the co-occurrence of the collinearity value between any two input variables and to detect the linear correlation between the independent/dependent (or input/target) variables^[29]. The statistically significant levels are based on the correlation coefficient which is used to evaluate the linear relationship between the two variables and is calculated by

$$p = \frac{PCC\sqrt{N-2}}{\sqrt{1-PCC^2}} \quad (2)$$

where the P -value is calculated via degrees of freedom (equal to $N-2$) and serves to assess whether the calculated correlation coefficient indicates the potential influence of the input/output variable. In this context, N denotes the total number of data points in the samples.

ML model selection and training

The data-driven approach in this study encompasses database development, ML model training, and the application of the trained model to screen N, O co-doped PCMs for supercapacitors. It also involves analyzing complex structure-activity relationships in characteristic variables and supercapacitor performance metrics. The ML work is executed using Python 3.9.7 software and the versions of all simulation tools are Jupyter Notebook 6.4.5, SHAP 0.41.0, and LGBM 4.3.0. Four classical ML prediction models such as LR, RF, XGBoost, and LGBM extensively were widely utilized to assess the intricate response relationships among various factors and the enhanced capacitance characteristics of of N, O co-doped PCMs. LR is one of the basic ML algorithms that is very efficient for linear datasets, small datasets, and simple relationships, but challenging for nonlinear datasets or highly complex data^[15]. RF is an algorithm based on ML that constructs and integrates decision trees, effectively addressing nonlinear regression issues among variables. It employs a technique referred to as “randomness” to generate a collection of decision trees, making it particularly well-suited for predictions based on small datasets^[30]. XGBoost is an efficient and scalable variant of the gradient boosting machine (GBM) algorithm. XGBoost has been widely used in data mining, recommendation systems and other fields due to its high efficiency, flexibility and portability^[23]. For XGBoost, to avoid over-fitting, an additional regularization term is added to the loss function to smooth the final learned weights. However, for data with nonlinear relationships, XGboost may not fit the model well, and it is difficult to adjust the parameters. LGBM is a framework for implementing the traditional gradient boosting decision tree (GBDT) algorithm, and an optimized algorithm based on XGBoost accelerates the training speed of the GBDT model without compromising accuracy. LGBM demonstrated higher accuracy, faster training speeds, a lower memory footprint, and the ability to handle large-scale data than traditional GBDT Tree models^[31].

In terms of ML models, various models might demonstrate superior or inferior performance based on the specific dataset under examination. While researchers frequently assess a model's fit to the training data as an indicator of its generalization capability, they also consider other factors beyond just the generalization error to gain a comprehensive understanding. To address this, the datasets were split into training and testing sets, where 80% was designated for training and 20% for testing.

ML models are capable of managing intricate relationships among dependent and independent variables, and they assess model quality through repeated cross-validation to prevent overfitting. The optimization process was performed using the k-fold cross-validation method with a 5-fold value to enhance the models' predictive performance.

Model evaluation

The quantitative assessment of different ML models to compare the prediction accuracy and quantify the prediction performance was carried out by utilizing the correlation coefficient (R^2) and root mean square error (RMSE) between the desired output and the estimated output provided by the ML model^[32]. The corresponding R^2 and RMSE values were given by

$$R^2 = 1 - \left(\frac{\sum_{i=1}^N (Y_i^e - Y_i^p)^2}{\sum_{i=1}^N (Y_i^e - Y_{ave}^e)^2} \right) \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_i^p - Y_i^e)^2}{N}} \quad (4)$$

where Y_i^e is the experimental actual value, and Y_i^p represents the predicted value. Y_{ave}^e stands for the mean of the experimental values.

Feature importance and partial correlation analysis

To investigate the complex effects of input feature factors on the SC of N, O co-doped PCMs, the feature importance of the model with good prediction performance was analyzed. SHAP plays a crucial role in enhancing the transparency and interpretability of this widely used ML methodology^[33]. In particular, the SHAP framework provides a systematic way to understand the complex mechanisms underlying the preferred model. It provides valuable insights into the impact of different input variables on model predictions. Hence, the SHAP was applied to calculate the Shapley value of each feature. SHAP values are provided based on how much each feature influences the resulting output. This revealed comprehensive relationships in how features contribute to the whole dataset between input variables and determined how they influence individual variables in the overall prediction. In addition, in order to further systematically display the relationship between variables and reflect the degree of influence of all features on prediction, we use single-feature and dual-feature interactive partial correlation graphs to perform partial correlation analysis on the preferred model^[34].

Verification experiment and deviation analysis

Preparation of N, O co-doped PCMs

Low-rank lignite was taken from the Inner Mongolia Autonomous Region (China) and demineralized with hydrofluoric acid (HF, about 20 wt.%) and hydrochloric acid (HCl, about 1 M) until the ash content was less than 1% to realize ash-free coal (named demineralization coal). Graphitic carbon nitride (g-C₃N₄) was first calcined by melamine as a precursor at 550 °C (heating rate: 5 °C·min⁻¹) for 2 h, then naturally cooled to room temperature to obtain the product. Other reagents include Potassium carbonate (K₂CO₃, Sinopharm

Chemical), potassium hydroxide [(KOH, $\geq 85\%$, Sinopharm Chemical), HF (40 wt.%), Shanghai Aladdin Reagent Co., Ltd.], HCl (36 wt.%, Sinopharm Chemical), polyvinylidene difluoride (PVDF, Shanghai Aladdin Reagent Co., Ltd.), N-methyl pyrrolidone (NMP, Shanghai Aladdin Reagent Co., Ltd.), and acetylene black (Shanghai Aladdin Reagent Co., Ltd.) were used directly without further purification. Deionized water was provided by our laboratory and used in all experiments. To further verify the accuracy and universality of ML prediction and consider the important influence of heteroatom dopants and activators in the synthesis of PCMs, N, O co-doped PCMs (named NOPC-x, x represents the number of different PCMs) were synthesized by regulating the mass ratio of carbon precursor, nitrogen dopant and catalyst based on NOPC-x synthesis procedure. Taking the pyrolysis temperature at 800 °C as an example, the pyrolysis procedure and conditions were the same as the above NOPC-x synthesis process, and the mass ratios of demineralized lignite, g-C₃N₄, and K₂CO₃ were set as 1:0:0, 1:1:0, 1:1:0.25, 1:1:0.5, 1:1:0.75, 1:1:1, 1:1:1.15, 1:1:2, respectively. It should be noted that “0” means without adding g-C₃N₄ or K₂CO₃. Finally, the resulting products were named NOPC-x, which x corresponds to numbers 1 to 8.

Characterization

N₂ adsorption-desorption was characterized using an ASAP 2020 PLUS HD88 analyzer (Micromeritics, USA) to obtain the SSA, PSD, and pore textures of all samples. The obtained sample (about 100 mg) was degassed at 160 °C for 8 h to remove impurities and water and then tested at 77 K with liquid nitrogen as adsorbate. The composition and structure characteristics of all samples were analyzed by X-ray photoelectron spectroscopy (XPS, Thermo Scientific K-Alpha) equipped with an Al K α radiation source. The binding energies were calibrated internally using the surface contamination C 1s binding energy of 284.8 eV.

Electrochemical measurement

For capacitance performance evaluation, galvanostatic charge-discharge (GCD) electrochemical measurements were analyzed using a three-electrode system at the CHI760E (Shanghai Chenhua) electrochemical workstation. In the three-electrode system, the Pt plate and Hg/HgO were used as counter and reference electrodes, respectively. The electrochemical performance of electrode materials was tested in a 6 M KOH aqueous electrolyte solution. In addition, working electrodes for testing were prepared before all electrochemical tests began. Specifically, the preparation of the working electrode was shown in the following process: a uniform homogeneous electrode slurry paste was obtained by mixing 80% active material with carbon black (10%) and PVDF (10%) with NMP as the solvent. The highly uniformity mixture was coated onto a carbon cloth (1 cm \times 1 cm) and dried at 80 °C overnight in a vacuum oven. The weight of the loaded active material on the dried electrode sheet was 1.28 mg. GCD curves were observed at different current densities (0.5 and 1.0 A·g⁻¹) when conducting constant current charge and discharge experiments under the voltage range of -1~0 V. The gravimetric SC (C, F·g⁻¹) was calculated in a three-electrode system using

$$C = (i_m t) / (\Delta V) \quad (5)$$

where the i_m (A·g⁻¹), ΔV (V), and t (s) represent the specific discharge current, the potential change, and the discharge time of the electroactive materials, respectively.

Under the condition of frequency range 0.01-100 kHz and amplitude 5 mV, the impedance curve data of the verified experimental sample were obtained by electrochemical AC impedance test, and the charge transfer resistance (R_{ct}) and equivalent series resistance (R_{es}) were calculated by fitting the equivalent circuit.

Deviation analysis

The characteristics of pores and microstructure features obtained from testing, the intensity of N/O doping and functional groups, and different CD values are used as values of input variables. On the basis of the optimal ML model, the SC performance was predicted under two CD values of 0.5 and 1 A·g⁻¹, respectively, and the predicted results were compared with the actual test results to verify the reliability of the prediction model. The deviation of the experimental and prediction results was further elucidated by a detailed analysis of experimental characterization data.

RESULTS AND DISCUSSION

Figure 1 provides a visual representation of the process of predicting SC through various ML models and traditional experimental validation. By likening it to a football match, the advantages of using ML for interpreting, predicting, and screening PCMs for supercapacitors are highlighted, showcasing its superiority over traditional experimental methods for material optimization. In general, the optimization of N, O co-doped PCMs to obtain high SC performance requires a complete experimental process; specifically, the football player (Experimental verification, EV) needs to go through the synthesis (S), characterization (C), and electrochemical testing (T) to achieve goals (means high SC performance). Three football player designs in reverse indicate that the continuous optimization of the three links can achieve the high SC of N, O co-doped PCMs, and also represents the time-consuming and laborious experiment process. The football player named ML, with the help of three players based on previous data studies: element composition (EC), pore characteristics (PF), and test environment (TC), can accelerate the screening and prediction of high-performance N, O co-doped PCMs. More deeply, we carried out ML-assisted study on intensively digging in the relationship between element composition, pore structure characteristics, operating conditions and SC performances. Through comprehensive interpretability analysis of ML, important variables influencing SC properties are identified and their impact range is determined. By analyzing the deviation of key values during experimental verification, accurate predictions of SC properties of N, O co-doped PCMs are made, facilitating precise material screening. Additionally, the accuracy and applicability of the ML model predictions are evaluated.

Descriptive statistical analysis

Microstructural properties, compositional properties, and testing conditions related to N, O co-doped PCMs were summarized using box-normal plots [Figure 2A-L and Supplementary Figure 1A-F]. The distribution of each variable aligns well with previous findings from meta-analysis [Supplementary Table 1]. Average and standard deviation (SD) values represent the dispersion of each feature, and minimum and maximum values describe the range of variables. Additionally, interquartile ranges are provided to offer supplementary location measures. Overall, all variables exhibited a normal distribution except for the PW. The voltage window predominantly ranges from 0.8 to 1.0 V [Supplementary Figure 1A], primarily due to the constraints imposed by the water decomposition voltage of 1.23 V^[35-37] in most water system supercapacitors, resulting in a narrow working window (≤ 1.0 V). CD [Figure 2B] is used as another parameter to test the SC, with values ranging between 0.5 and 50 A·g⁻¹. This parameter is mainly determined by the conventional test operation setting, and the SC of the corresponding test is adjusted to demonstrate the rate performance of the material. Pore structure characteristics play a crucial role in determining the structural properties of PCMs, which in turn affect their performance in the electrical double-layer capacitors (EDLC). The SSA and V_t of the prepared PCMs varied from 8 to 3,430 m²·g⁻¹ and 0.02 to 2.59 cm³·g⁻¹, respectively [Figure 2C and D]. This facilitates rapid ion transport by reducing the pathways for ion diffusion, and the abundant pore volume further aids in ion diffusion^[38]. The pore parameters [Supplementary Figure 1B-F] provide further insight into the pore structure characteristics. The PSD shown in Supplementary Figure 1B, ranging mainly between 1-5 nm, indicates a distribution skewed towards large micropores and small mesopores, which is beneficial for enhancing SC. S_{mic} [Supplementary Figure 1C] and



Figure 1. Schematic diagram of N, O co-doped PCMs search and theoretical predictions and experimental validation for SC. PCMs: Porous carbon materials; SC: specific capacitance.

V_{mic} [Supplementary Figure 1E], compared with total surface area and total pore volume, were prominently higher than the mesopore properties in the majority of the PCMs. The ratio of S_{mic}/SSA [Supplementary Figure 1D] typically ranged from 78% to 100%, while V_{mic}/V_t [Supplementary Figure 1F] varied from 62% to

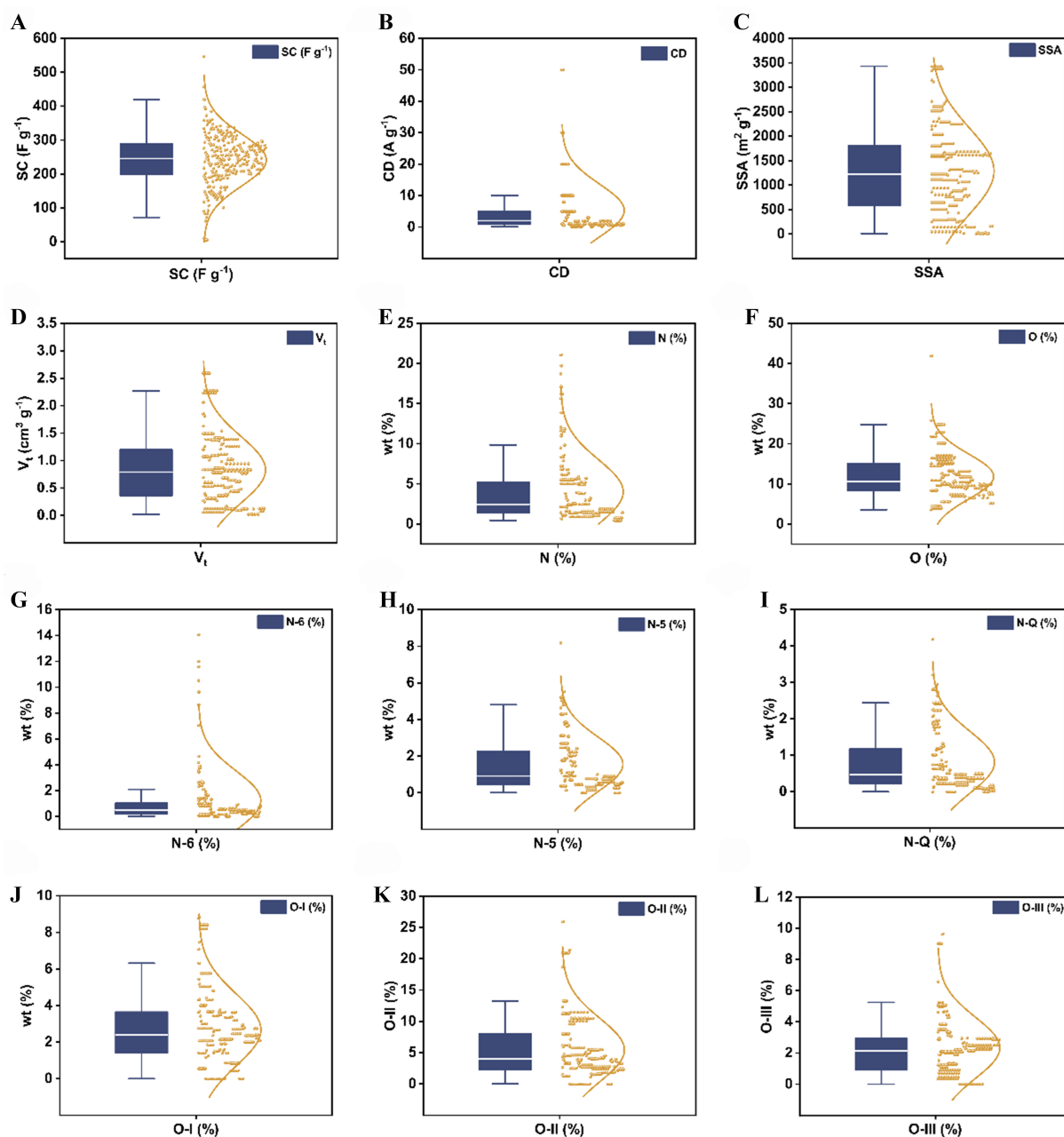


Figure 2. Box-normal plots of the data distributions for each parameter of collected data. (A) SC of PCMs; (B) Current density (CD) tested in a three-electrode system; (C) SSA of PCMs; (D) Total pore volume (V_t) of PCMs; (E and F) Contents of nitrogen (N), oxygen (O) in PCMs; (G-I) Contents of N-containing functional groups [pyridine nitrogen (N-6), pyrrole nitrogen (N-5), and graphitic nitrogen (N-Q)]; (J-L) Contents of O-containing functional groups [carbonyl-O group (C=O, O-I), hydroxyl-O/ether-O group (C-OH/C-O-C, O-II), and carboxyl-O group (-COOH, O-III)]. SC: Specific capacitance; PCMs: porous carbon materials; CD: current density; SSA: specific surface area.

98%. The pore structure characteristics, particularly the surface area and volume of micropores, play a significant role in increasing the porosity of heteroatom-doped PCMs' structures. The microporous structure is more conducive to active site generation and heteroatom doping. Analysis of the elemental composition revealed that N and O were present as heteroatom doping types in the PCMs. The O content ranged from 3.57% to 41.48% [Figure 2E], while the N content ranged from 0.43% to 21.06% [Figure 2F],

with an average N content of 4.00% and a maximum of 21.06%. This distribution reflects the current research focus on high N content doping. The incorporation of heteroatoms can modify the physical and chemical properties of porous carbon, impacting factors such as electrical conductivity and wettability, as well as inducing surface electroactive materials that influence capacity. Nitrogen doping in carbon materials can enhance electrical conductivity and properties. The presence of N or O functional groups in PCMs is crucial for pseudo-capacitance and overall supercapacitor characteristics. [Figure 2G-L](#) shows mean values for different nitrogen and oxygen groups (N-6 value of 1.22, N-5 value of 1.55, N-Q value of 0.81, O-I value of 2.67, O-II value of 5.43, O-III value of 2.33, respectively), suggesting potential incompatibilities among multiple dopants in PCMs. This highlights the need to consider various functional group factors in model development.

To analyze the dataset, PCC was computed between various features and capacitive performance [[Figure 3](#)]. The size of circles in [Figure 3](#) indicates correlation, with larger circles representing greater correlation. Notably, the results indicate that the correlations between feature characteristics and capacitance properties are relatively low ($PCC < 0.5$), suggesting that capacitance performance is influenced by complex processes involving multiple factors and their interactive effects, rather than being explained by a simple linear function. SC showed positive or inverse correlations with most independent variables, while CD exhibited a negative correlation with SC, with a correlation coefficient of -0.33 . Darker colors or larger circles in [Figure 3](#) denote stronger correlations between variables, whereas lighter colors indicate weaker correlations. [Figure 3](#) highlights that areas with darker colors are primarily found in regions associated with SSA, V_p , S_{mic} , and V_{mic} . Using all features to build a model may diminish model performance due to multicollinearity among variables. Nonetheless, taking into account the parameters that characterize pores offers a deeper insight into the differences between different pore materials. Therefore, they were preserved in the database despite a significant interrelation among them. Additionally, the relationship among other characteristic variables is minimal, suggesting that they operate independently. This improves the model's ability to generalize and reduces the occurrence of overfitting. Thus, six features (S_{mic}/SSA , V_{mic}/V_p , S_{mic} , V_p , SSA, PSD) with higher correlation values between porous structural features and SC were selected for modeling. In addition, most input parameters show a weak correlation with SC, which indicates that it is difficult to explore the quantitative influence of input parameters on the energy storage characteristics of PCMs through statistical analysis. Therefore, it is of great significance to build predictive ML models.

Model comparison and optimization

In order to ensure the accuracy of the model, we compared several models written in Python. Four traditional ML models (i.e., LR, RF, XGBoost, and LGBM) were adopted to accurately predict the capacitance of N, O co-doped PCM supercapacitors. The scatter plots in [Figure 4](#) display the predicted capacitance versus the real experimental capacitance of N, O co-doped PCM-based supercapacitors in the testing set. The predictive performance of each model was assessed using R^2 and RMSE metrics. Among the four models, LGBM showed the highest determination coefficient (R^2) of 0.92, with a relatively low RMSE of 2.79 [[Figure 4D](#)]. RF ranked second with an R^2 of 0.82 and RMSE of 37.08 [[Figure 4B](#)]. LR performed the least accurately with an R^2 of 0.44 [[Figure 4A](#)]. The predictive performance ranking of the ML models for N, O co-doped PCM supercapacitors is as follows: LGBM > RF > XGBoost > LR [[Figure 4](#)]. [Figure 4A-D](#) illustrates the disparities between the anticipated and experimental values across the four selected models utilized. The X-axis denotes the actual SC value, while the Y-axis illustrates the predicted values of SC calculated by each model. The closer the predicted value is to the actual value, the nearer the mark is to the bisecting line. Consequently, the gap between the mark and the bisecting line indicates the extent of error in the prediction. [Figure 4A](#) displays a more scattered distribution of points compared to other models, suggesting systematic errors in the prediction of LR model. The R^2 value for the optimal test set stands at 0.99, with the training set registering an R^2 value of 0.92, signifying superior prediction performance of the

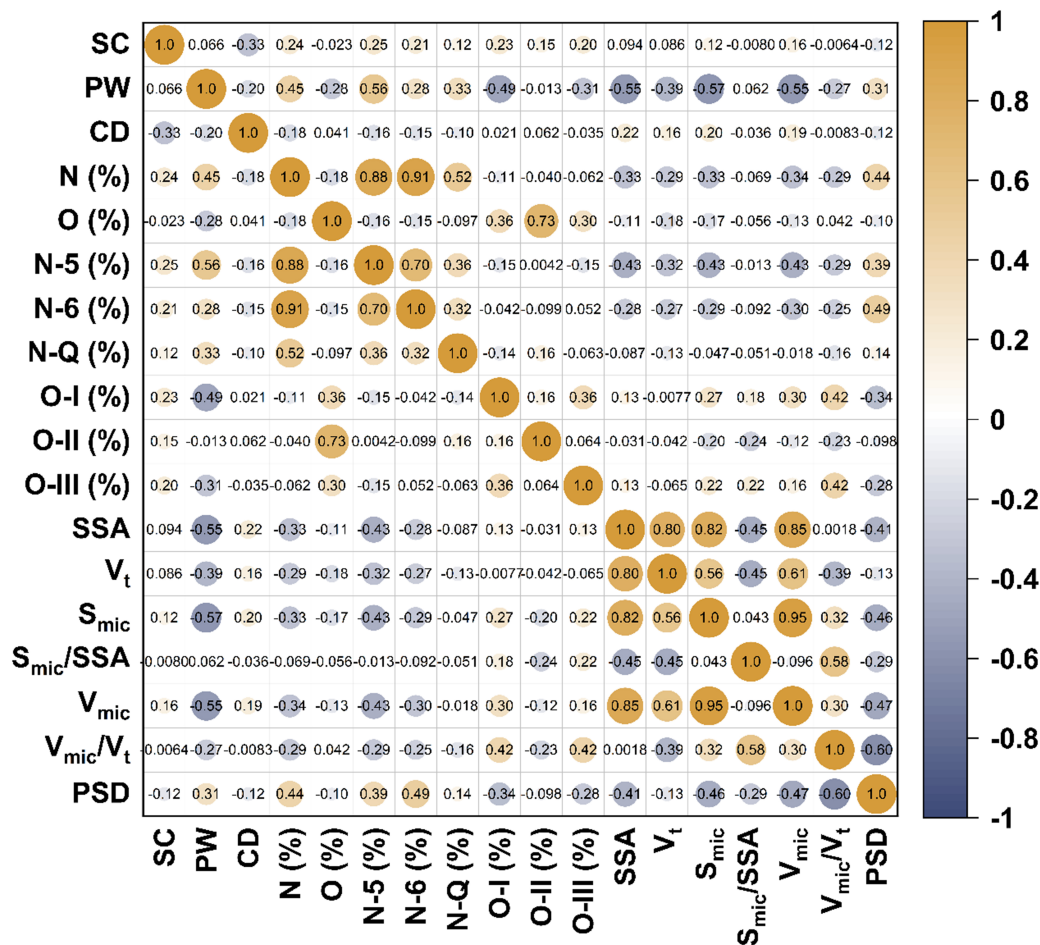


Figure 3. Heatmap analysis between all features and SC. The dimensions and hues of the circles in the illustration indicate the strength and direction of the correlation. They are distinguished by color codes: brown denotes positive effects, while dark blue indicates negative effects. SC: Specific capacitance.

LGBM model over other models. Overall, within this dataset, LGBM demonstrated the most accurate predictive capability during both training and testing phases, attributed to algorithmic variances and data range influences. From the point of view of ML model selection, LGBM gradually improves the predictive power of the model in the regression task by building a series of weak learners (usually decision trees). It optimizes the model by minimizing the loss function so that it can make accurate predictions of continuous target variables. Compared to traditional GBDT, LGBM uses a number of optimization techniques to improve performance. In addition, LGBM provides an early stop mechanism to avoid overfitting and improve training efficiency. Consequently, based on the model optimization outcomes, LGBM is designated as the preferred prediction model for subsequent simulation analyses.

Feature importance analysis

The importance of the characteristic variables was first considered, and their effects on the predictor variables were then analyzed. A higher importance value indicates a stronger influence on the output variable. In Figure 5A, we present the ranking of key factors determining the SC of PCMs. The results reveal that elemental composition and pore characteristics contribute to approximately 44.2% and 31.8% of PCMs, respectively, while 24.1% of the PCMs is attributed to the test condition [inset Figure 5A]. This implies that altering the test condition may have a limited impact on enhancing SC. The electrochemical test data in the

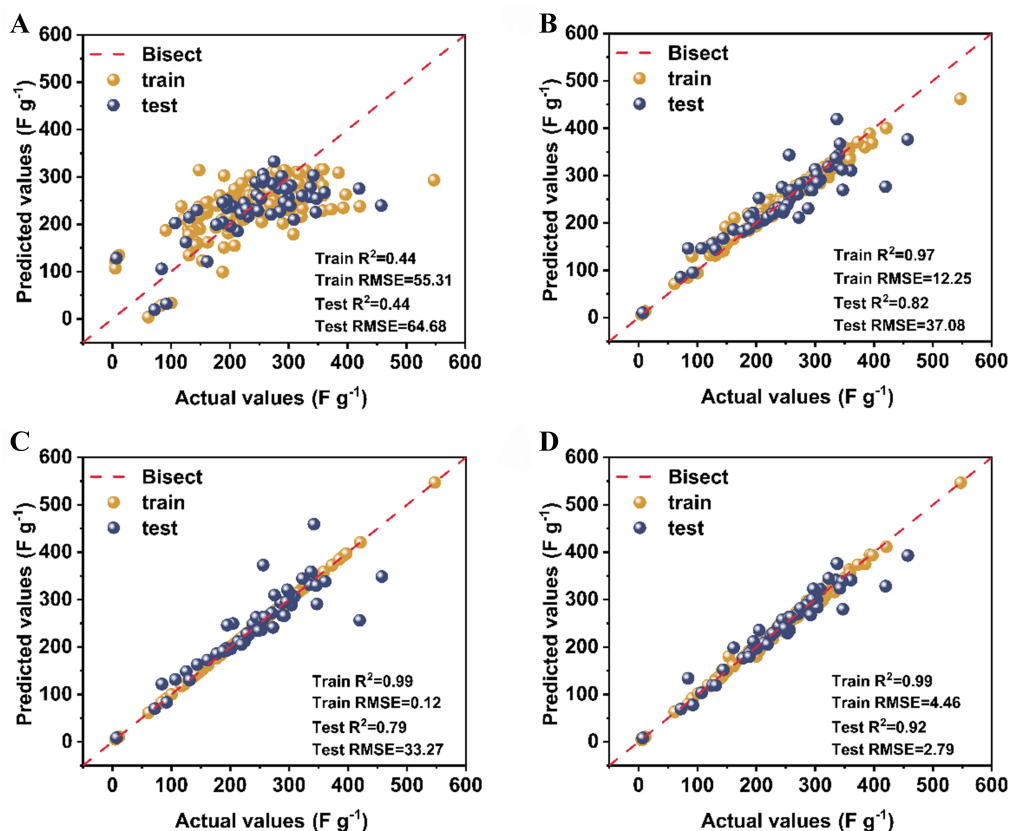


Figure 4. Modeling the performance of N, O co-doped PCMs for single-target prediction using (A) LR, (B) RF, (C) XGBoost, and (D) LGBM. PCMs: Porous carbon materials; LR: linear regression; RF: random forest; LGBM: light gradient boosting machine.

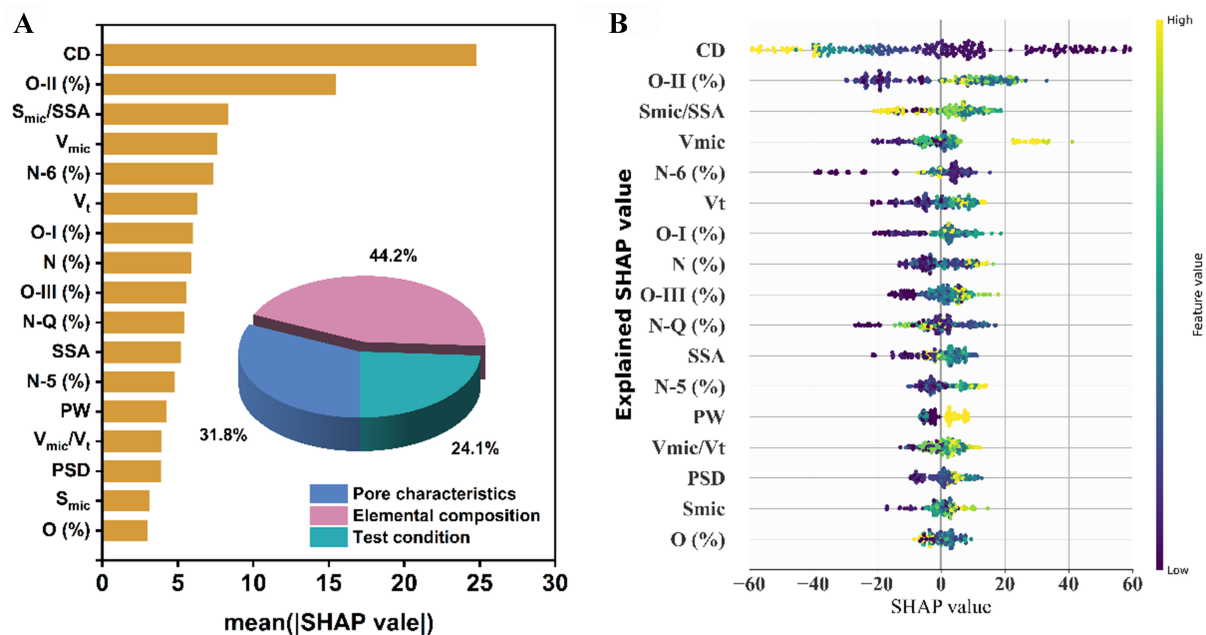


Figure 5. Feature importance analysis. (A) Order of relative contribution among variables; and (B) SHAP values of feature importance of targets (SC) obtained from LGBM model. SHAP: SHapley additive explanation; SC: specific capacitance; LGBM: light gradient boosting machine.

database are usually tested at different CDs, so the relationship between SC and CD can be constructed according to this. Therefore, CD ranks first in the ranking of model importance. Conversely, another test parameter, such as PW, exhibited a minor effect, largely due to the limited potential range of its aqueous supercapacitors. When excluding the influence of CD and PW data selection, the inherent characteristics of the raw porous material emerged as the most critical factors influencing capacitance prediction. Among these characteristics, O-II%, S_{mic}/SSA , V_{mic} , N-6%, V_t , O-I%, N%, O-III%, N-Q%, and SSA characteristics were the top ten variables significantly affecting SC performance. Feature importance was quantitatively analyzed to predict the relative contribution of each feature parameter in the LGBM model. The resulting SHAP values [Figure 5B], revealed that high CD values had a significant negative impact on SC prediction, whereas they positively impacted the SC prediction. The PW data showed a relatively narrow data range, leading to a more pronounced impact. Elemental properties such as oxygen and nitrogen and their type and content were identified as important factors. SHAP can be utilized to visualize the dependence of the model output (e.g., SC) on the value of each descriptor. The O-II% content of PCMs is positively related to SC [Figure 5B], particularly when it is below the SHAP value of 10, while negatively correlated to SC when the SHAP value of 10. Similarly, the distribution trends of O-I% and O-III% in PCMs follow a similar pattern in the SHAP analysis. These trends are due to the increase in active sites on carbon materials with the introduction of oxygen-containing functional groups, leading to enhanced charge transfer and improved interaction with electrolyte ions, thereby boosting wettability and pseudo-capacitance. However, excessively high levels of oxygen-containing functional groups can decrease SC due to factors such as blocked pore structure, increased ion transport resistance, and leakage current. The impact of oxygen-containing functional groups on PCMs is greater when the SSA and PSD are smaller. Adverse effects become more pronounced when oxygen-containing functional groups are present in excessive amounts^[39]. The pore characteristics of PCMs were identified as significant parameters in predicting SC. Among these characteristics, the S_{mic}/SSA and V_{mic} were found to be more influential than SSA, with the former two showing the highest importance after CD and O-II% [Figure 5A]. The SHAP values corresponding to these characteristics exhibited a notable positive correlation. The pore structure was observed to have an impact on SC, with higher values of V_{mic} or V_t associated with increased SC (as indicated by the yellow color in Figure 5B). Furthermore, both lower and higher values of S_{mic}/SSA were found to positively influence the ML model (LGBM), indirectly suggesting that a higher mesoporous or macro-porous surface area is linked to enhanced SC. The abundance of mesopores directly impacts electron transfer rates, leading to a shortened diffuse length and reduced diffusing resistance of ions in PCM supercapacitors. The increased number of N-doping plays a crucial role in enhancing their energy storage capabilities. When considering different percentages of N-doping, lower values generally have a negative impact on the ML model (LGBM), while higher values have a positive impact. However, the specific nitrogen-containing functional groups exhibit varying effects on the SC. For instance, both lower and higher values of N-6% or N-Q% negatively affect the ML model (LGBM), whereas overall N-5% shows a positive correlation with SC. The intricate impact of different types of nitrogen functional groups on the SC is attributed to the unpredictable conversion between nitrogen configurations during the synthesis of nitrogen-doped carbon materials, and the mixing of nitrogen species in the resulting PCMs, hindering the identification of electrochemically active nitrogen configurations for specific reactions. Most nitrogen-doped carbon materials undergo thermal treatment during synthesis, either through post-synthetic or in situ-doping methods. Thermal treatment above ~ 600 °C leads to nitrogen loss and uncontrollable conversion among nitrogen configurations, while treatment below ~ 600 °C yields poorly conductive nitrogen-doped carbon materials, limiting their electrochemical performance. Given that the pseudo-capacitance of PCMs is positively dependent on the N-5%, it can be inferred that N-5 species serve as highly active pseudocapacitive sites for nitrogen-doped carbon materials^[40]. Therefore, exploring and coordinating the relationship between the input features could enhance the SC of the PCMs.

Partial dependence analysis

The analysis of feature importance is utilized to demonstrate how a single variable affects the prediction target. One-dimensional partial dependence plots are then used to further explore how a single feature impacts the model prediction. Specifically, [Figure 6](#) and [Supplementary Figure 2](#) depict the one-dimensional partial dependence plots between each important feature and the target value SC. Within the range of 0-50 A·g⁻¹ [[Figure 6A](#)], CD exhibited a notable negative correlation with SC, aligning with the feature importance analysis. This is primarily due to higher CD leading to the attenuation of SC, which somewhat influenced the correlation analysis. PW showed a weak correlation with SC below 0.95 [[Figure 6B](#)], indicating that variations in the lower electric potential range did not lead to significant changes in SC. However, as PW approached 1 V, it exhibited a significant negative correlation with SC, further supporting the findings of the feature importance analysis. [Figure 6C-G](#) showed the effect of pore structure of PCMs on capacitance. The model in [Figure 6C](#) assumes that as the PSD value approaches 5, PCMs may achieve a higher SC during the SC prediction process. Conversely, for larger PSD values, this likelihood decreases.

Similarly, the V_{mic}/V_t in [Figure 6D](#) ranging from 0.5 to 0.8 and the S_{mic}/SSA in [Figure 6E](#) between 0.5 and 0.9 also indicate the potential for obtaining a high SC. Both V_t [[Figure 6F](#)] and SSA [[Figure 6G](#)] demonstrated similar positive effects with SC, with a significant increase in predicted SC as the input values increased. SSA values ranging from 500 to 2,500 m²·g⁻¹ and V_t values between 0.5 and 1.5 cm³·g⁻¹ were compared to V_{mic} values ranging from 0.5 to 1 cm³·g⁻¹ [[Supplementary Figure 2A](#)] and S_{mic} values between 500-2,000 m²·g⁻¹ [[Supplementary Figure 2B](#)]. These comparisons revealed a stronger correlation with SC, indicating that manipulating the micro/mesoporous SSA ratio and optimizing pore structure significantly influenced the energy storage characteristics of PCMs. In the analysis of doping characteristics of the original PCMs in the given parameters, a high feature importance is attributed to N%, while O% is ranked lower but still holds significance. Both variables exhibit a positive correlation with SC [[Figure 6E](#) and [H](#)], in line with prior studies. Notably, N% exhibits a stronger influence on SC than O%. This disparity can be attributed to the distinct impact of nitrogen dopants in the synthesis process on the energy storage of PCMs, while the O-doping primarily relies on the carbon precursor itself, resulting in a limited range of oxygen content due to pyrolysis loss.

The relationship between nitrogen content and SC is illustrated in [Figure 6H](#), showing that higher nitrogen content leads to increased SC. Conversely, [Figure 6I](#) demonstrates that excessive oxygen content can hinder SC enhancement. When considering N, O co-doped PCMs, it is essential to take into account not only the content of N or O but also types in which they exist after doping to accurately predict the SC of PCMs. Overall, higher input values of various functional groups result in higher SC predictions [[Figure 6J-O](#)]. However, within a lower range for different N or O containing functional groups, the SC value is predicted to fluctuate upward with the increase of each functional group value. This variation is mainly due to the significant effect of N or O doping on the energy storage characteristics of PCMs. Increasing such as N-5 [[Figure 6K](#)], O-I [[Figure 6M](#)], and O-II [[Figure 6N](#)] within a certain range effectively enhances the capacitance performance. The introduction of N imparts the carbon layer with an electron donor property, creating a rich electrochemical active site for the pseudocapacitive reaction. Notably, N-6 and N-5 species are identified as highly active pseudocapacitive sites for nitrogen-doped carbon materials^[40]. The analysis of N-Q partial correlations in [Figure 6L](#) confirms that traditionally the limited nitrogen doping level (1-3 wt%) usually hinders the exertion of positive effect from playing out and only stops at increasing the conductivity. Conversely, carbon materials with enriched N-species are of great desire to release the restricted properties, usually showing better electrochemical performance^[41]. To this end, it can be inferred that high nitrogen doping may be beneficial to the prediction of high SC. The partial correlation analysis of O-I [[Figure 6M](#)], O-II [[Figure 6N](#)], and O-III [[Figure 6O](#)] indicates that the correlation with SC varies with increasing content. This suggests that a moderate doping content of oxygen-containing functional groups is beneficial

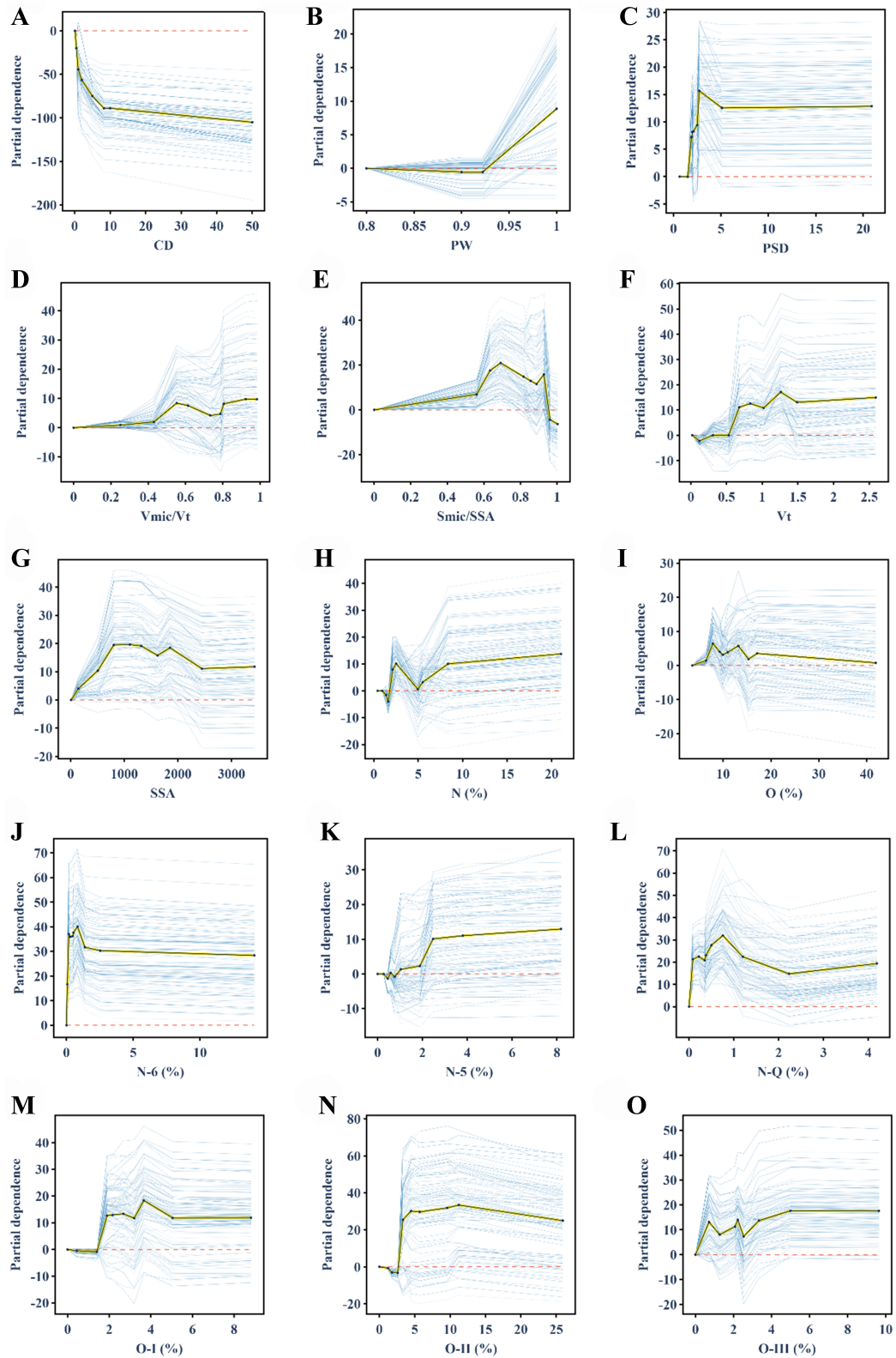


Figure 6. One-dimensional partial dependence plots of (A) CD, (B) PW, (C) PSD, (D) V_{mic}/V_t , (E) S_{mic}/SSA , (F) V_t , (G) SSA, (H) N%, (I) O%, (J) N-6, (K) N-5, (L) N-Q, (M) O-I%, (N) O-II%, and (O) O-III%. CD: Current density; PW: potential window; PSD: pore size distribution; SSA: specific surface area.

for SC^[42]. Additionally, **Figure 6M-O** demonstrates the positive effect of high oxygen doping characteristics or oxygen-rich functional groups of PCMs on predicting high SC.

In the actual preparation process of PCMs, achieving a balance or regulation of the intricate effects of pore characteristics and doping characteristics for SC can be challenging. The relationship between each variable of pore characteristics and doping characteristics has not been thoroughly explored. To gain a deeper understanding of the interactive influence between these features, two-dimensional partial correlograms were constructed for in-depth analysis [**Figure 7, Supplementary Figures 3 and 4**]. The interaction between pore structure-related features was analyzed, and the relevant results were presented in **Supplementary Figure 3**. In **Supplementary Figure 3**, positive effects were observed within a specific SSA distribution range of 500-2,500 m²·g⁻¹. Synthesizing carbon materials with an SSA of around 1,000 m²·g⁻¹ was recommended for high SC characteristics, particularly when V_t exceeded 0.68 cm³·g⁻¹. **Supplementary Figure 3B** indicated that a V_{mic} exceeding 0.82 cm³·g⁻¹, falling within the V_t range, was beneficial for boosting SC. Additionally, a PSD around 2.72 nm was suggested for enhancing capacitance [**Supplementary Figure 3C**]. **Supplementary Figure 3D** highlighted two distinct blue areas, each representing suitable pore volume proportions and PSDs. The corresponding V_{mic}/V_t ratios ranged from 0.55 to 0.61 or from 0.79 to 0.98, respectively. Notably, increasing V_{mic} and S_{mic} significantly impacted capacitance enhancement [**Supplementary Figure 3E**], with a more pronounced effect observed at high V_{mic} and S_{mic} values, likely due to a rise in adsorption sites. Similarly, the blue areas in **Supplementary Figure 3F** were significant, emphasizing the need for a comprehensive assessment when selecting the optimal interval. The total SC value reached its peak when the mass fractions of V_{mic}/V_t and S_{mic}/SSA varied from 0.55-0.61 and 0.63-0.82. Simultaneously, the SC value of V_{mic}/V_t fell within the maximum region in **Supplementary Figure 3F**. The analysis of interactions indicated that pore parameters significantly influenced capacitance prediction. It is advised to control SSA within the range of 500-2,500 m²·g⁻¹ and V_{mic} exceed 0.82 cm³·g⁻¹, while maintaining a desirable PSD around 2.72 nm during the preparation of PCMs. Additionally, adjusting the V_{mic}/V_t to 0.55-0.61 or 0.79-0.98 may be beneficial when the S_{mic}/SSA falls within the range of 0.56-0.82. Considering the positive impact of N or O content and species on the pseudo-capacitance of PCMs, it is crucial to elucidate the relationship between doping content and doping species to enhance SC. **Supplementary Figure 4A-F** illustrates the interaction effects of N content and N-containing functional group relative capacitance prediction. The predicted SC of PCMs exhibited the best effect when the N-6% fell between 0.47%-0.83% for low N-doping (N% < 5%) or range 0.18% to 0.83% for high N-doping (N% > 5%). Furthermore, the bidirectional increase of N-5 contents and N contents was observed to positively impact SC enhancement [**Supplementary Figure 4B**], with high nitrogen and high N-5 content correlating with higher capacitance prediction. This suggests that N-5 species function as highly active pseudocapacitive sites for nitrogen-doped carbon materials. Additionally, a moderate N-Q was found to contribute to achieving high SC [**Supplementary Figure 4C**], with the optimal N-Q distribution falling within the range of 0.38% to 1.2% for both low nitrogen (< 5%) and high nitrogen (> 5%) regions.

The interactive influence analysis of N-containing functional groups was shown in **Supplementary Figure 4D and F**. The optimal ranges for N-6%, N-5%, and N-Q% were found to be 0.47%-0.83%, 2.48%-8.18%, and 0.38%-1.2%, respectively, aligning with the insights from the one-dimensional partial dependence plots. Two-dimensional partial dependence diagrams illustrating the relationship between the content of O and O-containing species were presented in **Supplementary Figure 4G-L**. The mass ratio of O-I% ranged from 3.2% to 5.5% in **Supplementary Figure 4G**, indicating the best SC effect. **Supplementary Figure 4H and I** revealed that higher proportions of O-II% (5%-10%) and O-III% (5%-10%) were associated with increased SC in total O-doping. **Supplementary Figure 4J-L** depicted interactions among various O-containing functional groups. Detailed analysis showed that O-I% ranged from 2.16% to 5.05%, O-II% from 3.31% to

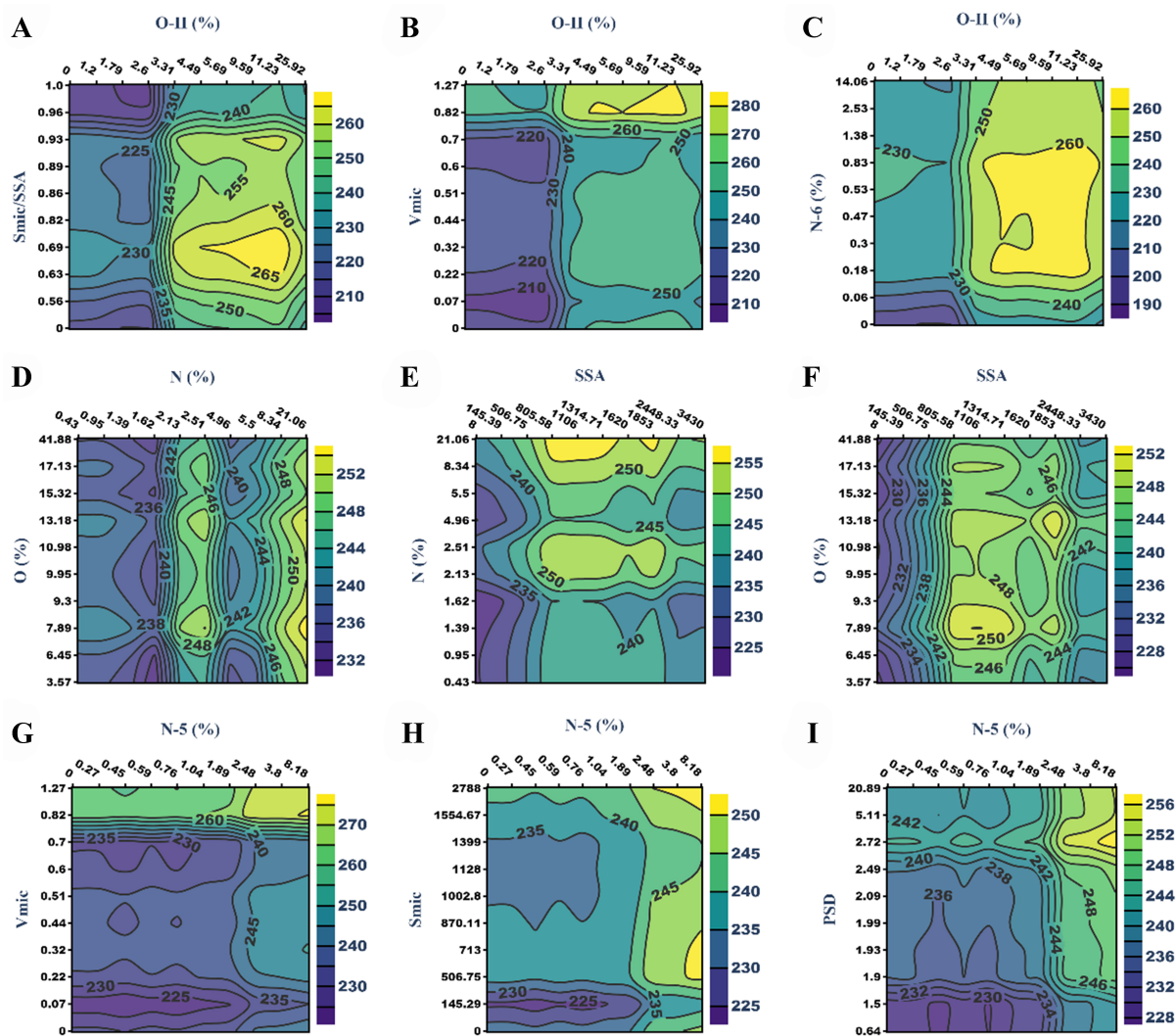


Figure 7. Two-dimensional partial dependence plots of (A) O-II% and S_{mic}/SSA , (B) O-II% and V_{mic} , (C) O-II% and N-6%, (D) N% and O%, (E) SSA and N%, (F) SSA and O%, (G) N-5% and V_{mic} , (H) N-5% and S_{mic} , (I) N-5% and PSD. SSA: Specific surface area; PSD: pore size distribution.

11.23%, and O-III% from 3.33% to 9.64%, suggesting potential for enhanced SC. O-II% had a more significant impact on SC compared to O-I% and O-III%, consistent with feature importance analysis results. While moderate doping and suitable pore structure are crucial for achieving high SC, the significant impact of doping content and species on SC necessitates consideration of their interactions with pore conditions. Therefore, we delved into the interaction between typical doping characteristics and pore-related features. Beginning with the top five features based on importance ranking, we conducted interactive analysis. Two-dimensional partial dependence diagrams depicting the relationship between O-II% with S_{mic}/SSA , V_{mic} , and N-6 % were illustrated in Figure 7A-C. With the increase of O-II% content, the landing area of S_{mic}/SSA ranges from 0.56 to 0.96 [Figure 7A], while the V_{mic} ranges from 0.82 to 1.27 $\text{cm}^3 \cdot \text{g}^{-1}$ [Figure 7B], indicating higher SC for PCMs. This is primarily attributed to the significant ratio of S_{mic}/SSA and V_{mic} , which play a major role in improving capacitive charge storage. The rise in O-II% content enhances the pseudo-capacitance active sites in reversible alkaline electrolytes, ultimately leading to an increase in SC. Figure 7C demonstrates that an increase in O-II% has a greater impact on SC compared to an increase in N-6%. When the O-II% content in PCMs is below 4.5%, incorporating a suitable amount of N-6% can enhance the SC.

However, with high O-II% content, maintaining N-6% within the range of 0.18%-1.38% is more likely to achieve high SC. The impact of O-II% outweighs that of N-6% in the feature importance analysis, a finding further supported by the two-dimensional interaction diagram. Additionally, the presence of doping elements must be carefully considered to fully elucidate the significant influence on SC enhancement. The interaction diagram of N% and O% in [Figure 7D](#) reveals that high N-doping content results in better SC than low N-doping within a moderate O-doping range. In scenarios of low and high N-doping, higher or lower O-doping levels do not necessarily translate to higher SC; however, O-doping in the range of 10%-15% or 6.45%-9.3% can contribute to SC improvement. This suggests that N-doping has a more positive impact than O-doping, aligning with the feature importance analysis results. The presence of N and O elements may participate in the Faraday reaction, and the combined effects of N-doping and O-doping are advantageous for enhancing capacitance. The drop point region of [Figure 7E](#) indicates that higher SC is observed at a SSA ranging from 500-2,500 $\text{m}^2\cdot\text{g}^{-1}$ with lower nitrogen (2.13%-2.51%) or higher nitrogen doping (5.5%-21.06%). This suggests that the combined effect of N-doping and SSA is beneficial for enhancing SC. When the SSA is below 500 $\text{m}^2\cdot\text{g}^{-1}$, changes in N-doping content have a significant impact on SC, while variations in O-doping content within the range shown in [Figure 7F](#) have a relatively minor influence on the SC. Moreover, within the SSA range of 800-2,500 $\text{m}^2\cdot\text{g}^{-1}$, an optimal level of O-doping can lead to higher SC. The N-5 functional group, a representative species of N-doping, is considered highly efficient for enhancing SC. Consequently, the interaction between pyrrole nitrogen and pore characteristics was investigated. Interactions between N-5% and V_{mic} , S_{mic} , and PSD are illustrated in [Figure 7H-J](#), revealing that PCMs with higher V_{mic} generally yield higher SC. Additionally, increasing N-5% further enhances SC. In [Figure 7H](#), a larger S_{mic} , a higher N-5%, also corresponds to a higher SC prediction. When the PSD falls within the range of 0.64-5.11 nm and N-5% content is below 2.48%, increasing PS and doping content collectively promote SC enhancement. Conversely, when the N-5 doping level exceeds 2.48%, PSD exerts a greater influence on SC improvement. In summary, through feature importance analysis, univariate dependency graph and two-dimensional interactive influence analysis, key features for preparing N, O co-doped PCMs with high SC were identified. These include $S_{\text{mic}}/\text{SSA}$ (0.63-0.82 or 0.89-0.96), V_{mic}/V_t (0.55-0.61 or 0.79-0.98), SSA (500-2,500 $\text{m}^2\cdot\text{g}^{-1}$), PSD (0.64-5.11 nm), lower N% (2.13%-2.51%) or higher N-doping (5.5%-21.06%), O% (below 15%), N-5% (2.48%-8.18%), and O-II% (3.31%-11.23%).

Experimental validation, prediction and deviation analysis

Through feature importance analysis, univariate dependence graph and two-dimensional interaction analysis, the key feature range for preparing N, O co-doped PCMs with high SC performance was determined. Ideal values of SSA (1,965 $\text{m}^2\cdot\text{g}^{-1}$), $S_{\text{mic}}/\text{SSA}$ (0.74), V_{mic}/V_t (0.91), PSD (2.11 nm) were achieved based on LGBM by considering multi-physiochemical and test features. Also, the obtained values of N%, O%, N-5%, and O-II% were 17.57 at%, 14.5 at%, 5.34 at%, and 3.98 at%, respectively. On this basis, combined with the actual sample test conditions, considering the three-electrode system under the condition of 6 M KOH, the LGBM-based model predicts that the maximum SC of N, O co-doped PCMs is 469 $\text{F}\cdot\text{g}^{-1}$ at 0.5 $\text{A}\cdot\text{g}^{-1}$ in a PW of 1.0 V. The capacitance value obtained by the optimization process is somewhat different from the value predicted by the previous ML ^[14,26,42], which may be related to the database collected and the model used. Furthermore, the rationality of the constructed model and the selection of features require further validation through experiments to ensure accurate prediction. Validation experiments on eight NOPC samples at 0.5 $\text{A}\cdot\text{g}^{-1}$ and at 1.0 $\text{A}\cdot\text{g}^{-1}$ displayed GCD profiles in [Figure 8A](#) and [B](#). The GCD curves exhibit a linear shape with slight deviations, indicating a blend of electrical double-layer and pseudo-capacitor energy storage mechanisms attributed to porous and N, O co-doped pseudo-capacitor sites. The SC values of the prepared NOPC samples, as predicted by the LGBM model [[Figure 8C](#)]. Test results reveal actual SC values ranging from 228.4 and 357 $\text{F}\cdot\text{g}^{-1}$ at 0.5 $\text{A}\cdot\text{g}^{-1}$ and from 142.6 to 330.3 $\text{F}\cdot\text{g}^{-1}$ at 1.0 $\text{A}\cdot\text{g}^{-1}$. A comparison between predicted and actual values indicates some deviation, particularly as CD increases, leading to fluctuations in prediction accuracy. The prediction accuracy of SC

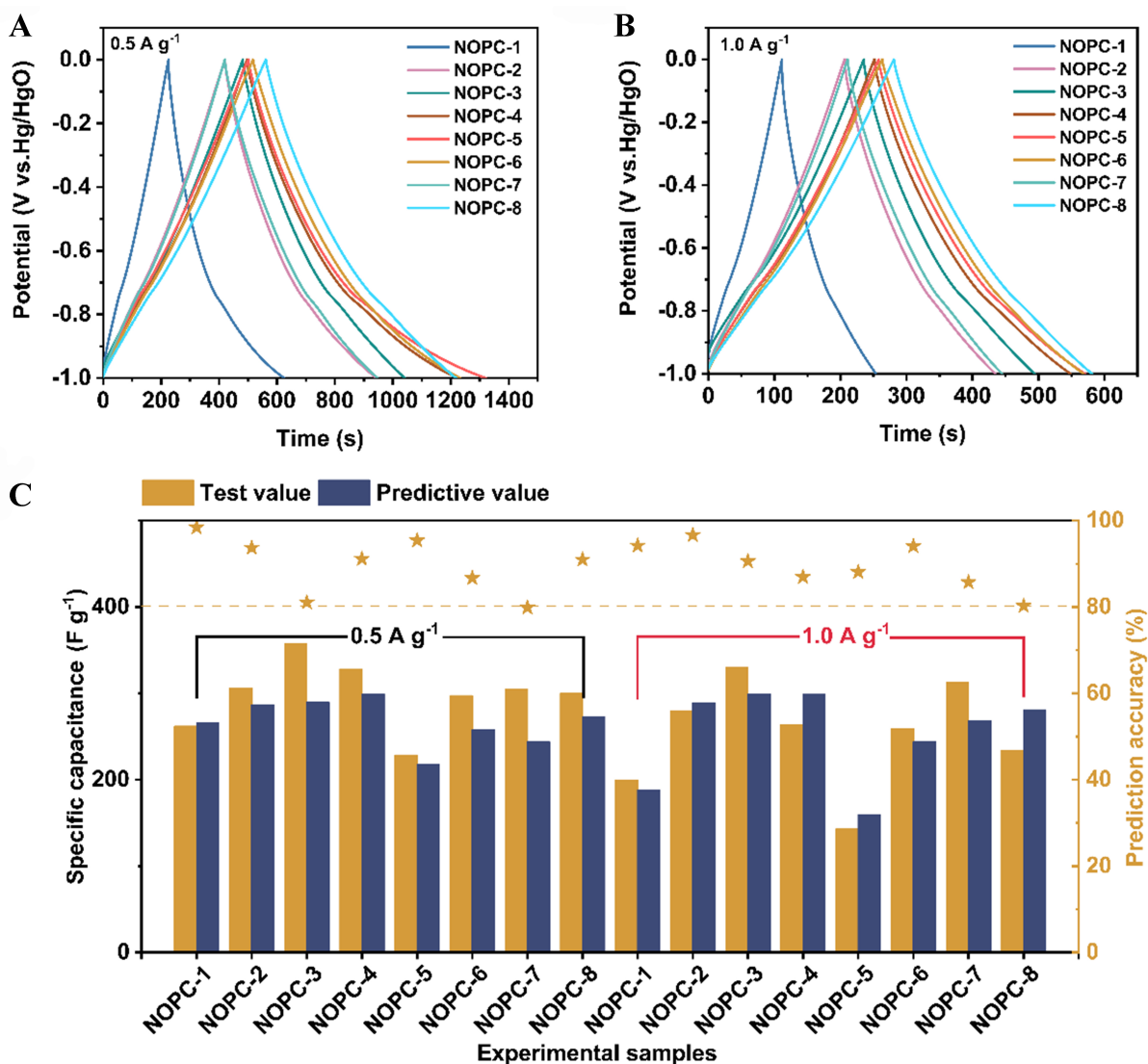


Figure 8. Capacitance performance (SC) assessed through a three-electrode configuration. (A) GCD profiles of eight NOPC samples at 0.5 A g⁻¹ and (B) at 1.0 A g⁻¹ in electrochemical experiments; (C) Test and predicted values of SC about eight NOPC samples in verification experiments and prediction accuracy. SC: Specific capacitance; GCD: galvanostatic charge-discharge; NOPC: nitrogen, oxygen co-doped porous carbon materials..

ranges between 80% and 98%, differing from the R^2 (test set). This discrepancy is primarily attributed to significant deviations in SSA, V_p , N% or O% of the prepared samples from the predicted eigenvalue range, directly impacting the energy storage performance of N, O co-doped PCMs and resulting in biased results. For this purpose, Brunauer-Emmett-Teller (BET) and XPS characterization tests were used to analyze the pore structure and composition structure of PCMs (NOPC samples), respectively. The nitrogen adsorption-desorption isotherms were shown in [Supplementary Figure 5A](#). The sharp increase of adsorption volume ($P/P_0 < 0.05$) and hysteresis loop ($P/P_0 = 0.45-0.9$) indicated that there were abundant micropores and mesopores in each sample synthesized by verification experiment, revealing the hierarchical porous structure of PCMs. The PSD curves estimated using density functional theory (DFT) confirmed the significant micropore distribution and some mesopores [[Supplementary Figure 5B and C](#)]. Among the samples from NOPC-1 to NOPC-8, the micropores of PCMs are mainly concentrated at about 0.6, 0.9 and 1.3 nm, accompanied by abundant mesopores in the range of 2-4 nm. More detailed pore

parameters such as SSA, S_{mic} , PSD, V_t , S_{mic}/SSA , V_{mic}/V_t , and V_t of NOPC samples were summarized in [Supplementary Table 2](#). In all synthesized NOPC samples, S_{mic}/SSA distribution was 0.25-0.932; V_{mic}/V_t distribution was 0.111-0.969; SSA ranged from 12 to 1,192 $m^2 \cdot g^{-1}$; PSD was in the narrow range of 1.79-3.00 nm. The specific elemental compositions and their chemical states or functional group species of the experimentally verified samples (NOPC samples) were revealed using XPS tests. In the full XPS spectrum of NOPC samples, there are three absolute peaks at 284.8, 399.7 and 532.7 eV, which belong to the peaks of C, N, and O, respectively [[Supplementary Figure 6A](#)], indicating the successful co-doping of N and O elements. Further, the doping forms of N and O elements in the carbon skeleton were shown in [Supplementary Figure 6B-D](#). Among them, The nitrogen doping forms are as follows, containing pyridine nitrogen (N-6, 398.69 eV), pyrrole nitrogen (N-5, 400.03 eV) and graphitic nitrogen (N-Q, 401.28 eV), respectively. O-doped forms are as follows: carbonyl-O group (C=O, O-I, 531.04 eV), hydroxyl-O/ ether-O group (C-OH/C-O-C, O-II, 532.49 eV), and carboxyl-O group (-COOH, O-III, 533.51 eV)^[43]. The specific elemental compositions and their species doped by N and O were summarized in [Supplementary Table 3](#). In particular, the doped content of N elements ranges from 1.82% to 17.9%, the doped content of O elements ranges from 3.04% to 9.46%, the content of N-5 species ranges from 0.51% to 5.49%, and the distribution of O-II species ranges from 0.83% to 3.43% in the actual synthesized NOPC samples. Combined with the analysis of capacitance performance in [Figure 8C](#), the lowest capacitance performance of NOPC-1 in the actual experimental sample may be related to its low N and O doping (the N and O element in this sample is derived from the carbon precursor low-rank lignite itself). The nitrogen doping content of other samples is above 5%, and the oxygen doping content is about 5% or higher, which means higher capacitance performance. Lowest SSA and V_t in NOPC-3 may be related to high N-doping, high O-doping and low addition of activators without sufficient activation. Data deviation analysis between experimental features and predicted important features such as S_{mic}/SSA , V_{mic}/V_t , SSA, PSD, N%, O%, N-5%, and O-II% was presented in [Figure 9](#). Further application of local displayed shapes to explain certain prediction deviations is shown in [Supplementary Figure 7](#). Various characteristic indicators of NOPC-1 in [Figure 9](#), such as O-II%, SSA, N%, and N-5%, deviated from the predicted range of important features, yet achieved prediction accuracies of 98.42% ($0.5 A \cdot g^{-1}$) and 94.2% ($1.0 A \cdot g^{-1}$), respectively, with slight deviations from the laboratory values compared to the predicted capacitance. Local SHAP analysis showed that the prediction bias of NOPC-1 is negatively affected by O-II% and the change of CD has a positive effect on the prediction. At different current densities, the features have different effects on the prediction. This is mainly due to the fact that the collected data is not completely satisfactory, and may be due to some special data outside the model prediction range resulting in inaccurate forecasts. Notably, the S_{mic}/SSA , V_{mic}/V_t , and SSA eigenvalues of NOPC-3 exhibited significant deviations from the predicted range, leading to a larger deviation compared to the capacitance prediction. Local SHAP analysis showed that S_{mic}/SSA and V_{mic}/V_t had positive effects, so S_{mic}/SSA and V_{mic}/V_t may be the main causes of prediction bias in the actual NOPC-3. Prediction discrepancies for NOPC-6 and NOPC-7 were primarily attributed to the low O-II% content, as supported by the feature importance ranking of O-II%. In contrast to NOPC-6, the prediction accuracy of NOPC-7 improved with increasing CD ($A \cdot g^{-1}$), largely due to the influence of SSA, which enhances ion and charge adsorption and conversion. In the local shape analysis, O-II% showed a major negative effect in NOPC-6, and O-II% and SSA showed a major negative effect as the CD ($A \cdot g^{-1}$) changed. Further analysis indicated an average testing accuracy of approximately 90% for all NOPC samples, which is basically consistent with the model R^2 value. However, the unexpectedly high accuracy and minimum difference in the NOPC-1 sample suggest that the overall composition of the PCMs also plays a role, emphasizing the need to consider the effects of the PCMs themselves, synthesis parameters, and electrochemical test parameters when integrating the prediction range of important feature analysis. Nyquist curves, equivalent circuit diagrams and fitting curves of all NOPC samples were further shown in [Supplementary Figure 8A-D](#). All curves are semicircular in the high-frequency region, oblique in the mid-frequency region, and almost vertical in the low-frequency region. The R_{ct} and R_{es} calculated by fitting were presented in [Supplementary Table 4](#). Specifically, the R_{ct}

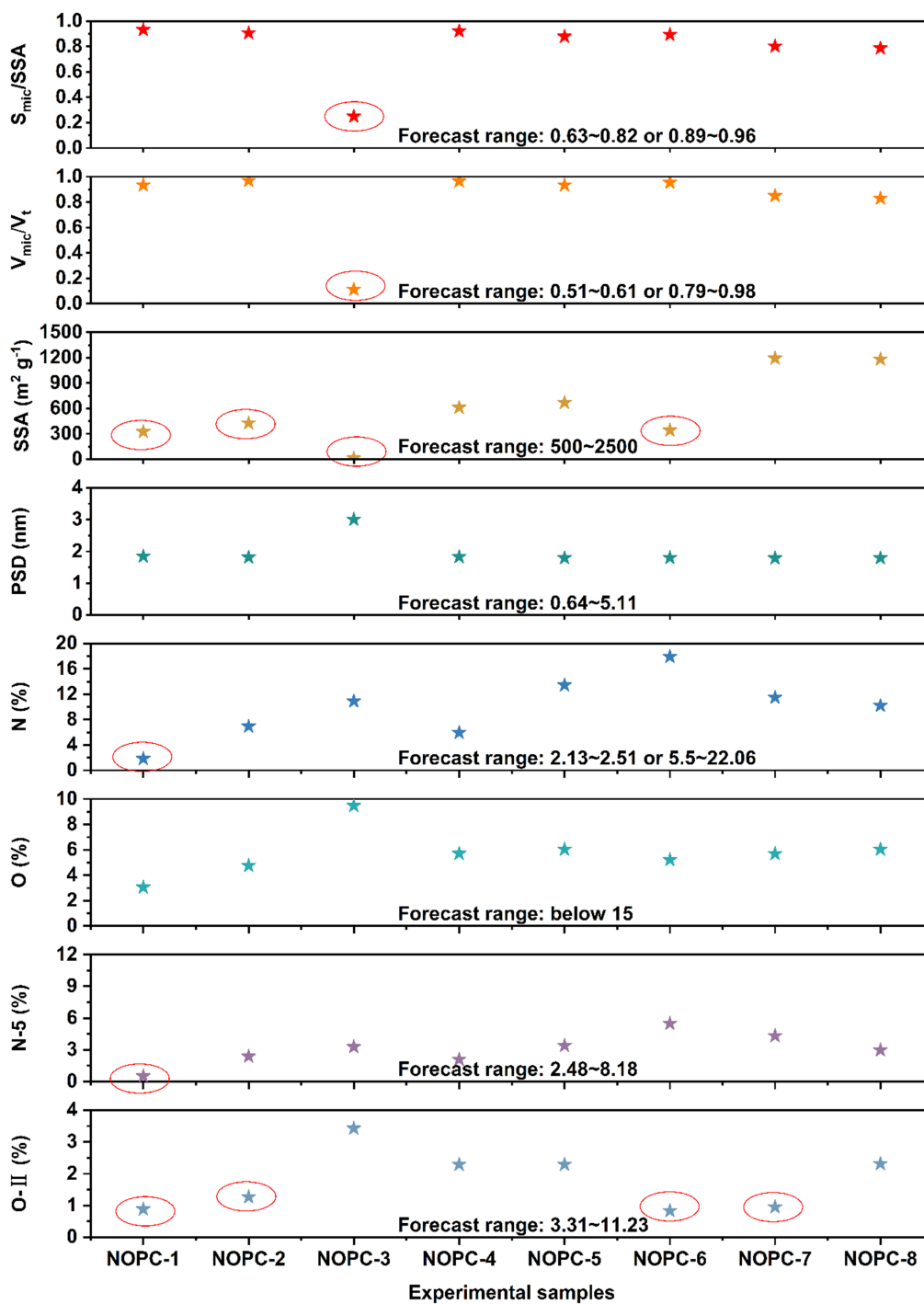


Figure 9. Deviation analysis of experimental feature indexes in the range of prediction key features (S_{mic}/SSA , V_{mic}/V_t , SSA, PSD, N%, O%, N-5%, and O-II%). SSA: Specific surface area; PSD: pore size distribution.

ranges from 0.003 to 2.62 Ω , and the Res ranges from 0.58 to 3.37 Ω . NOPC-1 showed that the highest Rct and Res may be related to its underdeveloped pore structure resulting in its lowest SC performance. All other samples have small semicircles and low Rct, indicating faster ion diffusion and charge transfer dynamics. The Res from NOPC-2 to NOPC-8 all showed low values, indicating that the prepared materials

have good electrical conductivity, which is related to the introduction of graphitized carbon in the prepared samples [Supplementary Table 3], and may be beneficial to the increase of SC. Considering that the overall conductivity of the carbon material was found to be the result of interactions among multiple factors, texture and surface chemistry were probably the most relevant, and the conductivity as selected features in ML (Rct and Res properties based on electrochemical impedance tests) had an impact on the next stage of the prediction process^[44,45]. Validation experimental results demonstrated that the model could be further optimized by incorporating appropriate synthetic parameters and more comprehensive electrochemical test features, but accuracy enhancement would require additional ML processes. Compared with traditional experimental methods, the limited amount of data may not be able to fully cover all situations. This shows that ML models should still be used in conjunction with practical applications, and underscores the importance of collecting high-quality data in predicting applications of high-performance porous carbon-based energy storage materials.

Challenges and prospects

Compared with traditional experiments, the ML model adopted in this study and the proposed method can save experiment cost and time, while providing accurate predictions. Our work provides new insights into the screening and prediction of N, O co-doped PCMs, and this can help accelerate the development of high-performance carbon materials.

However, there are some limitations to this study:

- (1) Any model has its own advantages and inherent limitations for predicting the target. The LGBM regression model learns the relationship between features and target variables from the data when trained, so the performance of the model may suffer when encountering test data with a significantly different distribution than the training data. LGBM cannot predict the target value beyond the feature range of training data, and is sensitive to skewed data. It is also affected by feature engineering and preprocessing.
- (2) Some special values in the literature dataset have some influence on the prediction performance of the model because they are out of the data distribution range, which leads to the deviation of the prediction results.
- (3) The characteristics used in this study mainly involve the elemental composition, pore structure and capacitance test conditions of N and O co-doped PCMs, while other characteristics or electrochemical tests such as conductivity, defects and electrode preparation process parameters may affect capacitance through different mechanisms, and the pyrolysis and activation processes are not considered. In the actual material preparation process, there are complex process combinations that affect the composition characteristics, so it is necessary to consider different process parameters and further refine the characteristic parameters to predict their effects.

Future research should collect characteristic parameters based on experimental synthesis, characterization and test results as much as possible to expand data quantity and improve data quality. In addition, it is necessary to establish a reliable prediction model on the basis of the existing analysis results and try to develop a new model to further improve the science, accuracy, and practicability of the model. For other feature parameters collected, the input parameters should be adjusted in time, and the ML prediction model should be built by using similar processes. Meanwhile, the model should be updated and optimized in real time, and finally establish a system for carbon-based energy storage materials' prediction.

CONCLUSIONS

In summary, four ML models, LR, RF, XGBoost, and LGBM, were utilized to predict the SC property of N, O co-doped PCMs. The prediction results indicated that the LGBM model outperformed the other three algorithm models, achieving an R^2 value of 0.92 on the prediction set. Through SHAP analysis, one-dimensional partial dependence analysis and two-dimensional interaction diagram analysis, it was determined that higher values of S_{mic}/SSA (0.63-0.82 or 0.89-0.96), V_{mic}/V_t (0.55-0.61 or 0.79-0.98), larger SSA (500-2,500 $m^2 \cdot g^{-1}$), PSD (0.64-5.11 nm), lower N% (2.13%-2.51%) or higher N-doping (5.5%-21.06%), O% (below 15%), N-5% (2.48%-8.18%), and O-II% (3.31%-11.23%) were found to be advantageous for enhancing the SC of N, O co-doped PCMs. Experimental validation using NOPC samples was conducted to assess the applicability and the accuracy of the preferred LGBM model. Comparison and analysis of the deviation based on the fall-point region of key features and experimental feature value facilitate the prediction and screening of N, O co-doped PCMs. Functional importance (SHAP, partial correlation analysis) based on the best LGBR model provides reasonable explanations for these variables. All in all, the ML model developed in this study offers cost and time savings while maintaining accurate predictions. The combination of ML model predictions, experimental verification, and deviation analysis of falling data proposed in this study presents a novel approach to enhancing the accuracy of SC prediction and accurate screening of N, O co-doped PCMs.

DECLARATIONS

Authors' contributions

Conceived the idea and designed the project: Liu H, Wang Y

Performed data analysis and interpretation: Liu H, Cui Z

Supervised the project: Liu H, Qiao Z, An X

Drafted the manuscript: Liu H

Revised and finalized the manuscript: Liu H, Cui Z, Qiao Z, An X, Wang Y

All authors read and approved the final manuscript.

Availability of data and materials

[Supplementary Materials](#) are available from the *Journal of Materials Informatics* or the authors.

Financial support and sponsorship

The work was supported by the National Natural Science Foundation of China (No. 52371231) and the Key R&D Program of Shanxi Province (No.202302040201008).

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2024.

REFERENCES

1. Pohlmann S. Metrics and methods for moving from research to innovation in energy storage. *Nat Commun* 2022;13:1538. DOI PubMed PMC
2. Pandya DJ, Muthu Pandian P, Kumar I, et al. Supercapacitors: review of materials and fabrication methods. *Mater Today Proc* 2023;In Press. DOI
3. Liu X, Lyu D, Merlet C, et al. Structural disorder determines capacitance in nanoporous carbons. *Science* 2024;384:321-5. DOI PubMed
4. Li J, Xia Z, Wang X, et al. Distinguished roles of nitrogen-doped Sp² and Sp³ hybridized carbon on extraordinary supercapacitance in acidic aqueous electrolyte. *Adv Mater* 2024;36:e2310422. DOI PubMed
5. Yang M, Zhou Z. Recent breakthroughs in supercapacitors boosted by nitrogen-rich porous carbon materials. *Adv Sci* 2017;4:1600408. DOI PubMed PMC
6. Yang J, Su F, Liu T, Zheng X. Heteroatoms co-doped multi-level porous carbon as electrode material for supercapacitors with ultra-long cycle life and high energy density. *Diam Relat Mater* 2024;141:110693. DOI
7. Zhang Y, Wang Y, Feng Y, Liu W, Chen F. Amorphous nickel-cobalt phosphate nanosheets optimized by a micron-sized copper particle array with ultrahigh specific capacitance for asymmetric supercapacitors. *J Alloys Compd* 2024;990:174393. DOI
8. Patil S, Bhosale A, Kundale S, Dongale T, Vanalakar S. Enhancing capacitance performance of functional group assisted carbon quantum dots derived from turmeric plant waste. *Carbon Trends* 2024;15:100370. DOI
9. Hajibaba S, Gholipour S, Pourjafarabadi M, et al. Electrochemical sulfur-doping as an efficient method for capacitance enhancement in carbon-based supercapacitors. *J Energy Storage* 2024;79:110044. DOI
10. Wood KN, O'Hayre R, Pylypenko S. Recent progress on nitrogen/carbon structures designed for use in energy and sustainability applications. *Energy Environ Sci* 2014;7:1212-49. DOI
11. Dong K, Wang J, Guo F, et al. Facile synthesis of O and P co-doped hierarchical porous carbon nanosheets from biomass for high-performance supercapacitors. *Diam Relat Mater* 2023;140:110531. DOI
12. Zheng Y, Chen K, Jiang K, Zhang F, Zhu G, Xu H. Progress of synthetic strategies and properties of heteroatoms-doped (N, P, S, O) carbon materials for supercapacitors. *J Energy Storage* 2022;56:105995. DOI
13. Lu Y, Liang J, Deng S, et al. Hypercrosslinked polymers enabled micropore-dominant N, S Co-doped porous carbon for ultrafast electron/ion transport supercapacitors. *Nano Energy* 2019;65:103993. DOI
14. Zhou M, Vassallo A, Wu J. Data-driven approach to understanding the *in-operando* performance of heteroatom-doped carbon electrodes. *ACS Appl Energy Mater* 2020;3:5993-6000. DOI
15. Jha S, Yen M, Salinas YS, Palmer E, Villafuerte J, Liang H. Machine learning-assisted materials development and device management in batteries and supercapacitors: performance comparison and challenges. *J Mater Chem A* 2023;11:3904-36. DOI
16. Thakkar P, Khatris S, Dobariya D, Patel D, Dey B, Singh AK. Advances in materials and machine learning techniques for energy storage devices: a comprehensive review. *J Energy Storage* 2024;81:110452. DOI
17. Rahimi M, Abbaspour-Fard MH, Rohani A. A multi-data-driven procedure towards a comprehensive understanding of the activated carbon electrodes performance (using for supercapacitor) employing ANN technique. *Renew Energy* 2021;180:980-92. DOI
18. Jha S, Bandyopadhyay S, Mehta S, et al. Data-driven predictive electrochemical behavior of lignin-based supercapacitors *via* machine learning. *Energy Fuels* 2022;36:1052-62. DOI
19. Pan R, Gu M, Wu J. Data-driven optimization of carbon electrodes for aqueous supercapacitors. *J Chem Eng Data* 2024. DOI
20. Su H, Lin S, Deng S, Lian C, Shang Y, Liu H. Predicting the capacitance of carbon-based electric double layer capacitors by machine learning. *Nanoscale Adv* 2019;1:2162-6. DOI PubMed PMC
21. Adekoya GJ, Adekoya OC, Ugo UK, Sadiku ER, Hamam Y, Ray SS. A mini-review of artificial intelligence techniques for predicting the performance of supercapacitors. *Mater Today Proc* 2022;62:S184-8. DOI
22. Yang H, Fang L, Yuan Z, et al. Machine learning guided 3D printing of carbon microlattices with customized performance for supercapacitive energy storage. *Carbon* 2023;201:408-14. DOI
23. Pan K, Liu Q, Zhu L, et al. Integrated data mining for prediction of specific capacitance of porous carbon materials for flexible energy storage devices. *J Energy Storage* 2023;73:109072. DOI
24. Liu P, Wen Y, Huang L, et al. An emerging machine learning strategy for the assisted-design of high-performance supercapacitor materials by mining the relationship between capacitance and structural features of porous carbon. *J Electroanal Chem* 2021;899:115684. DOI
25. Saad AG, Emad-Eldeen A, Tawfik WZ, El-Deen AG. Data-driven machine learning approach for predicting the capacitance of graphene-based supercapacitor electrodes. *J Energy Storage* 2022;55:105411. DOI
26. Rahimi M, Abbaspour-Fard MH, Rohani A. Synergetic effect of N/O functional groups and microstructures of activated carbon on supercapacitor performance by machine learning. *J Power Sources* 2022;521:230968. DOI
27. Sun Y, Sun P, Jia J, et al. Machine learning in clarifying complex relationships: Biochar preparation procedures and capacitance characteristics. *Chem Eng J* 2024;485:149975. DOI
28. Chenwittayakhachon A, Jitapunkul K, Nakpalad B, et al. Machine learning approach to understanding the 'synergistic' pseudocapacitive effects of heteroatom doped graphene. *2D Mater* 2023;10:025003. DOI
29. Yang X, Yuan C, He S, Jiang D, Cao B, Wang S. Machine learning prediction of specific capacitance in biomass derived carbon materials: effects of activation and biochar characteristics. *Fuel* 2023;331:125718. DOI

30. Liu X, Ji D, Jin X, Quintano V, Joshi R. Machine learning assisted chemical characterization to investigate the temperature-dependent supercapacitance using Co-rGO electrodes. *Carbon* 2023;214:118342. DOI
31. Kim M, Kang S, Gyu Park H, Park K, Min K. Maximizing the energy density and stability of Ni-rich layered cathode materials with multivalent dopants via machine learning. *Chem Eng J* 2023;452:139254. DOI
32. Sawant V, Deshmukh R, Awati C. Machine learning techniques for prediction of capacitance and remaining useful life of supercapacitors: a comprehensive review. *J Energy Chem* 2023;77:438-51. DOI
33. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: 31st Conference on Neural Information Processing Systems (NIPS 2017); Long Beach, USA. 2017. Available from: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf. [Last accessed on 23 Oct 2024]
34. Zhang Y, Feng Y, Ren Z, et al. Tree-based machine learning model for visualizing complex relationships between biochar properties and anaerobic digestion. *Bioresour Technol* 2023;374:128746. DOI PubMed
35. Kim H, Hong J, Park KY, Kim H, Kim SW, Kang K. Aqueous rechargeable Li and Na ion batteries. *Chem Rev* 2014;114:11788-827. DOI PubMed
36. Tan J, Liu J. Electrolyte engineering toward high-voltage aqueous energy storage devices. *Energy Environ Mater* 2021;4:302-6. DOI
37. Hwang JY, El-Kady MF, Li M, et al. Boosting the capacitance and voltage of aqueous supercapacitors via redox charge contribution from both electrode and electrolyte. *Nano Today* 2017;15:15-25. DOI
38. Hu J, Zhao C, Si Y, et al. Chitosan-derived large surface area porous carbon via microphase separation engineering of pore-regulation and nitrogen-doping coupling for high-performance supercapacitors. *Renew Energy* 2024;228:120598. DOI
39. Qiu C, Jiang L, Gao Y, Sheng L. Effects of oxygen-containing functional groups on carbon materials in supercapacitors: a review. *Mater Design* 2023;230:111952. DOI
40. Tian K, Wang J, Cao L, et al. Single-site pyrrolic-nitrogen-doped sp²-hybridized carbon materials and their pseudocapacitance. *Nat Commun* 2020;11:3884. DOI PubMed PMC
41. Chen T, Luo L, Luo L, et al. High energy density supercapacitors with hierarchical nitrogen-doped porous carbon as active material obtained from bio-waste. *Renew Energy* 2021;175:760-9. DOI
42. Wang T, Pan R, Martins ML, et al. Machine-learning-assisted material discovery of oxygen-rich highly porous carbon active materials for aqueous supercapacitors. *Nat Commun* 2023;14:4607. DOI PubMed PMC
43. Zhang S, Zhang Q, Ma R, et al. Boosting the capacitive performance by constructing O, N co-doped hierarchical porous structure in carbon for supercapacitor. *J Energy Storage* 2024;82:110569. DOI
44. Barroso-Bogeat A, Alexandre-Franco M, Fernández-González C, Macías-García A, Gómez-Serrano V. Temperature dependence of the electrical conductivity of activated carbons prepared from vine shoots by physical and chemical activation methods. *Micropor Mesopor Mat* 2015;209:90-8. DOI
45. Tawfik WZ, Mohammad SN, Rahouma KH, Salama GM, Tammam E. Machine learning models for capacitance prediction of porous carbon-based supercapacitor electrodes. *Phys Scr* 2024;99:026001. DOI

