Supporting Information

Prediction of discharge performance of hard carbon materials for high-performance sodium-ion batteries based on machine learning

Tianshuang Qi [a], Xiong Zhang [a, b]*, Kai Xiong [a], Haiping Yang [a], Shihong Zhang [a], Hanping Chen [a]

[a] *School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan 430074, China.*

[b] *Research Institute of Huazhong University of Science and Technology in Shenzhen, 518000 Shenzhen, PR China*

**Corresponding information: Zhang Xiong; 1037 Luoyu Road, Wuhan 430074, Hubei, China;*

Tel.: +086-27-87542417; Fax: +086 027-87545526; E-mail address: zhangxiong107@163.com

Table A.1

The correlation between input features and sodium storage mechanisms.

Mechanistic correlation	Input feature	Feature interpretation
Chemical structure (adsorption mechanism)	The contents of P, S, N, and O	Heteroatoms introduce extrinsic defects, provide active sites for adsorption, and enhance the surface adsorption capacity.
Microcrystalline structure (insertion mechanism)	d_{002}, La, Lc and n	d_{002}: Graphite layer spacing, affecting the dynamics and capacity of ion embedding. La, Lc: microcrystal size (transverse/longitudinal), which determines the continuity of the interlayer channel. n: indicates the number of stacked layers, which affects the storage space between layers.
Graphitization and defects (Integrated mechanism)	I_D/I_G	I_D/I_G: Raman spectral parameters, reflecting the degree of graphitization (high ratio = low graphitization, high defect density), the synergistic effect of associated ion adsorption and embedding.
Pore structure (filling mechanism)	D, V_{Total} and SSA	D: The average pore size affects the rapid adsorption and diffusion of sodium ions. V_{Total}: Total pore volume, which affects the total amount of sodium stored. SSA: Specific surface area provides more surface adsorption sites than surface area.

Cyclic factor definition

To more objectively reflect the cycle performance of hard carbon anodes, efforts have been made to identify the relationship between cycle performance and discharge current density. However, the existing literature has not adequately addressed a comprehensive and quantitative analysis of this relationship. The capacity retention rates of several sets of lithium-ion batteries at various discharge rates have been collected to address this gap. The specific data are presented in Table A.2.

Table A.2

Capacity retention rate of lithium-ion battery under different conditions at different discharge rates. (Data from the article DOI:10.16009/j.cnki.cn13-1295/tq.2020.05.002)

Discharge rate/C	0.2	0.5	1	2	3
Capacity retention rate /% (Group 1)	99.73	99.35	97.42	89.60	76.05
Capacity retention rate /% (Group 2)	99.78	99.41	97.77	91.48	79.18
Capacity retention rate /% (Group 3)	99.86	99.41	98.18	93.48	85.35
Capacity retention rate /% (Group 4)	99.95	99.82	98.36	93.15	89.41
Capacity retention rate /% (Group 5)	99.70	99.27	94.01	82.68	72.51

It is not difficult to find that the capacity retention rate of lithium-ion batteries decreases with the increase in discharge density. To find the relationship between the two, we try to fit with $P = a + b \times i^n$, and use R^2 (equation 12) to reflect the fitting effect. P is the capacity retention rate. Detailed results are shown in Table A.3.

Table A.3

The determination coefficient R^2 calculated by fitting under different n values.

n	Function expression	R^2 (Group 1)	R^2 (Group 2)	R^2 (Group 3)	R^2 (Group 4)	R^2 (Group 5)
1	$P = a + b \times i$	0.94274	0.93342	0.94951	0.98242	0.99191
2	$P = a + b \times i^2$	0.99954	0.99851	0.99990	0.95970	0.96163
3	$P = a + b \times i^3$	0.98245	0.98814	0.97936	0.88528	0.88759
4	$P = a + b \times i^4$	0.94826	0.95873	0.94369	0.81626	0.82101

In summary, the fitting effect of $P = a + b \times i^2$ is better, and the R^2 calculated by $P = a + b \times i^2$ remains above 0.95. Therefore, the relationship between lithium-ion battery capacity retention rate and discharge density may be set as $P = a + b \times i^2, b < 0$; Since the discharge rate is directly proportional to the discharge density[1], in addition, sodium and lithium have similar electrochemical mechanisms in the same main group of the periodic table. Therefore, the relationship between the capacity retention rate of the sodium ion battery and the discharge density can also be expressed by $P = a + b \times i^2$. For convenience, let $P = a - b \times i^2, b > 0$; where P is the capacity retention rate after 100 cycles, and i is the discharge density (mA/g); Because when $i = 0$, it means that the discharge current is zero, at this time, the capacity remains unchanged. Then $P = 100$, so $a = 100$; So, $b = (100 - P)/i^2$, we intend to define b as

the cyclic factor, the larger b, the worse the cycle performance.

Lithium-ion batteries and sodium-ion batteries share many core similarities in the mechanisms of storing alkali metal ions. From the physicochemical basis of charge storage, both Li^+ and Na^+ participate in electrochemical reactions with a +1 valence state, exhibiting single-charge characteristics and achieving charge transfer through redox reactions. Moreover, both complete storage through an intercalation/deintercalation mechanism, relying on the reversible structure of the electrode material, which can either intercalate into the lattice of the electrode material or be achieved through surface adsorption. In terms of electrode material design, despite the difference in ionic radii (Li^+: 0.076 nm, Na^+: 0.102 nm), both require adjustable interlayer spacing to accommodate ion diffusion and need to avoid structural collapse caused by repeated intercalation/deintercalation during cycling. Regarding dynamic interfacial behavior, alkali metal ions will react with the electrolyte on the anode surface to form an SEI film, the stability of which directly affects the cycle life. The storage process is also limited by the interfacial ion transport rate and the bulk diffusion barrier, manifesting as similar charge/discharge plateau characteristics. Furthermore, in the competitive pathways of energy storage, in addition to the intercalation mechanism, both systems may involve surface pseudocapacitive behavior, exhibiting energy storage characteristics with coexisting multiple mechanisms. Therefore, it is reasonable for us to draw on this dependency relationship from lithium-ion batteries.

It should be emphasized that the preceding analysis does not imply that a definitive relationship between the capacity retention rate of hard carbon materials and current density has been established. The capacity retention rate is influenced by a complex interplay of factors, including the structure of both the cathode and anode materials, the nature of the electrolyte, and the operating temperature. Consequently, the relationship between current density and capacity retention rate can be likened to a black box model, where the internal causal mechanisms cannot be directly observed or inferred from external data alone. Nevertheless, compared with the previous scenario where current density was not considered at all, this approach still holds some significance. It can reflect that under the condition of the same capacity retention rate, the greater the

current density, the worse the cycle performance of the battery is generally.

Calculation principle of D_{Na^+}

D_{Na^+} is calculated by the following formula[2]:

$$D_{Na^+} = \frac{4}{\pi\tau}\left(\frac{m_B V_M}{M_B S}\right)^2\left(\frac{\Delta E_S}{\Delta E_\tau}\right)^2 \quad \text{(A.1)}$$

Where m_B is the mass of the electrode active material; M_B is the molar mass of the material; V_M is the molar volume of the material; M_B/V_M is the density of the hard carbon material; S is the geometric area of the electrode sheet; τ is the current pulse time; ΔE_S and ΔE_τ can be obtained from the GITT curves, where ΔE_S refers to the value of the change in the material electrode potential after applying a current for some time other than the voltage drop portion, and ΔE_τ refers to the change in the is the difference between the potential after the current is applied and relaxation stabilized and the potential before the current is applied.

ML models

The selection of machine learning models should be based on the characteristics of the data and the requirements of the task. Different algorithms have their own strengths in terms of performance, efficiency, and applicability. Ensemble learning has become mainstream due to its powerful generalization ability:(1) XGBoost: It iteratively adds decision trees to correct prediction residuals. By introducing L1/L2 regularization to constrain the complexity of tree structures and combining pre-sorting with parallel computing techniques, XGBoost significantly speeds up training while reducing the risk of overfitting. It is the preferred choice for handling large-scale data [3]. (2) GBDT: As a traditional gradient boosting algorithm, GBDT lacks regularization and is sensitive to noise. However, its simple logic of residual fitting still performs well on small-scale nonlinear data [4]. (3) CatBoost: It proposes innovative solutions for categorical features. By using target variable statistics and ordered boosting strategies to

automatically encode discrete features, and employing symmetric binary trees to reduce overfitting, CatBoost shows significant advantages in scenarios with a large number of categorical variables [5]. (4) Random Forest: It constructs multiple independent decision trees and integrates their results to reduce model variance. By selecting the optimal split points from a random subset of features at each node, Random Forest balances accuracy and generalization ability[6]. (5) Extra Trees: It further introduces randomness in split points, sacrificing some accuracy to significantly increase training speed and enhance resistance to overfitting. This makes it suitable for scenarios with high real-time requirements. (6) SVR: By mapping data to a high-dimensional space using kernel functions and finding the maximum-margin hyperplane, SVR can effectively capture the intrinsic patterns of high-dimensional sparse data. However, the choice of kernel function and parameter tuning significantly impact performance, and computational complexity increases exponentially with the amount of data [7]. (7) KNN: As a typical lazy learning algorithm, KNN relies on local similarity for prediction. Although it requires no training and is intuitive, it suffers from the "curse of dimensionality" in high-dimensional spaces due to the failure of distance calculations. Additionally, the prediction phase requires scanning the entire dataset, leading to significant efficiency bottlenecks [8]. (8) MLP: Through multiple layers of nonlinear transformations, MLP can fit complex patterns and theoretically approximate any function relationship. However, it demands a large amount of data and computational resources, and is prone to performance fluctuations due to gradient vanishing or overfitting. Techniques such as Dropout and batch normalization are often used to optimize its performance. (9) ElasticNet: By combining Lasso (L1) and Ridge (L2) regularization, ElasticNet balances feature sparsity and multicollinearity handling capabilities through the adjustment of parameters. It is particularly suitable for linear regression problems with high-dimensional and strongly correlated features, but its performance is highly dependent on hyperparameter tuning. (10) AdaBoost: By dynamically adjusting sample weights to force subsequent base learners to focus on difficult-to-classify samples, AdaBoost can significantly improve the performance of weak classifiers. However, its sensitivity to label noise and outliers limits its application

in noisy data scenarios [9]. (11) Stacking Ensemble: By integrating the predictions of multiple heterogeneous base models (such as SVM, tree models, and neural networks) using a meta-learner, Stacking can theoretically break through the ceiling effect of a single model. However, its complex multi-stage training process is highly prone to overfitting, and computational costs increase exponentially. The design of base model diversity and meta-model complexity must be carefully considered in practical applications. (12) Bagging is an ensemble learning method that involves generating multiple distinct sub-training sets by sampling the training data with replacement. It then trains multiple independent base learners (typically decision trees) and aggregates their outputs through averaging to produce the final prediction. The primary objective of Bagging is to reduce the model's variance and enhance its generalization capability.

In summary, XGBoost and Random Forest, with their robust performance and high degree of automation, have become the "default options" for general scenarios. CatBoost demonstrates unique value in tasks with dense categorical features. SVR and ElasticNet provide specialized solutions for high-dimensional sparse or strongly linearly correlated data. Extra Trees and KNN meet the needs for lightweight deployment and simple prototype verification, respectively. Meanwhile, Stacking and MLP play key roles in complex systems with abundant resources and a pursuit of high precision. The core of model selection lies in balancing computational efficiency, interpretability, data characteristics, and business tolerance for errors, rather than blindly pursuing algorithmic complexity.

Table A.4
The data structure of the original (unfilled) data in this study.

Feature	Unit	Average value	Standard deviation	Minimum	25th	50th	75th	Maximum	△	Q3+1.5IQR	Q1-1.5IQR
P	at%	0.17	0.88	0	0	0	0	14.36	14.36	0	0
S	at%	0.47	2.22	0	0	0	0	22.083	22.083	0	0
N	at%	1.85	4.24	0	0	0	1.62	34.4	34.4	4.05	-2.43
O	at%	6.40	4.20	0	3.62	5.66	8.6	29.71	29.71	16.07	-3.85

d_{002}	Å	3.81	0.20	3.36	3.70	3.81	3.91	4.96	1.60	4.225	3.385
La	Å	48.11	19.22	11.65	35.91	47.13	59.50	148	136.35	94.885	0.525
Lc	Å	16.71	7.68	3.50	11.62	15.60	20.50	88.10	84.60	33.82	-1.7
n	-	8.74	2.15	0.88	3.04	4.02	5.25	26.22	25.34	8.565	-0.275
I_D/I_G	-	1.33	0.47	0.168	1.01	1.15	1.61	2.90	2.732	2.51	0.11
D	nm	7.94	9.99	0.206	2.71	5.16	8.11	84.95	84.744	16.21	-5.39
V_{Total}	cm^3/g	0.22716	0.406	0.00014	0.017	0.073	0.249	3.0531	3.05296	0.597	-0.331
SSA	m^2/g	247.52	471.29	0.23	10.36	50.28	296.17	3993	3992.77	724.885	-418.355
lg(SSA)	-	1.70	0.89	-0.64	1.02	1.70	2.47	3.60	4.24	4.645	-1.155
Cyclic factor	% g^2/mA^2	0.011888	0.023693	0.00004	0.000238	0.002568	0.008438	0.163475	0.16347476	0.020738	-0.012062
Rate factor	h	0.22	0.214	0.01	0.1	0.16	0.26	2.22	2.21	0.5	-0.14
ICE	%	64.16	17.63	14.07	52.8	66.37	78	98	83.93	115.8	15
Capacity	mAh/g	235.5	92.17	2.4	171.1	231.7	286.0	740.2	737.8	458.35	-1.25
Working Plateau	v	0.40	0.21	0.09	0.24	0.35	0.5	1.28	1.19	0.89	-0.15
$-LgD_{Na^+}$	-	10.09	1.88	4.37	9.05	9.85	11.34	15.5	11.13	14.78	5.62

△ Represents the difference between the maximum and minimum values; Q1 is the value at 25% in the data; Q3 is the value at 75% in the data; IQR is the quartile distance and IQR= Q3-Q1.

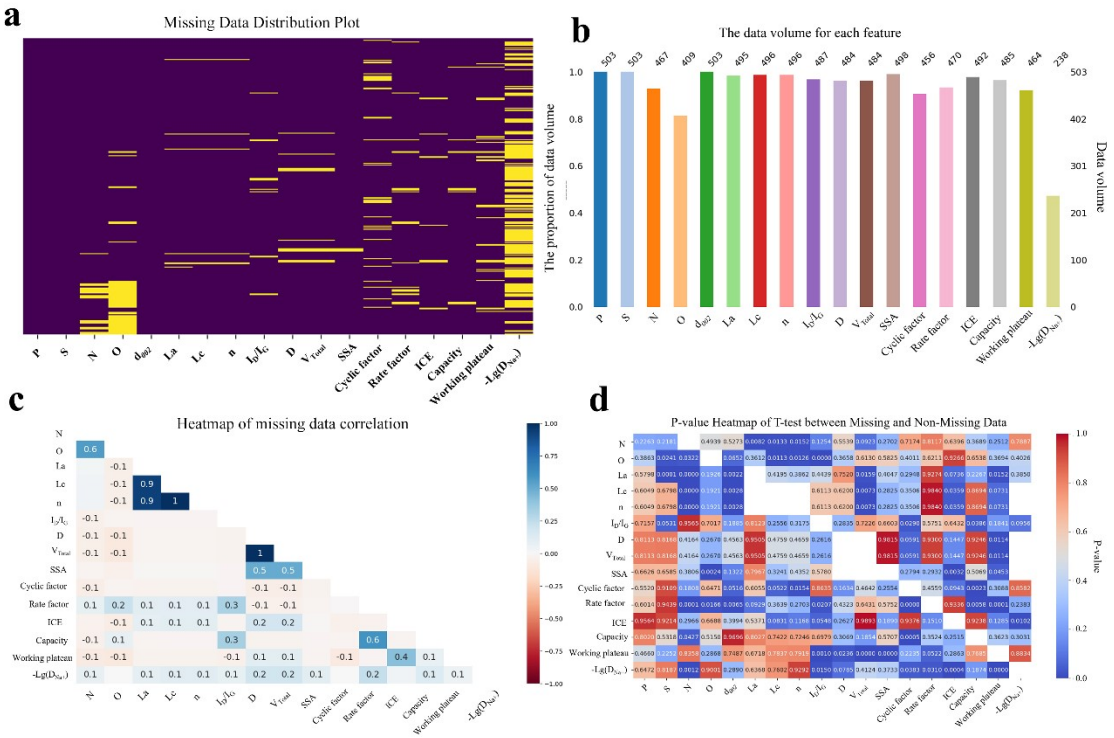

Fig A.1 The impact of missing values on the dataset in this study. (a) Distribution of missing values. Yellow blanks in the figure indicate missing data. (b) The number and proportion of missing values for each variable. (c) Heatmap of missing data correlation. A correlation coefficient closer to +1 indicates a stronger positive correlation; a correlation coefficient closer to -1 indicates a stronger negative correlation; if the correlation coefficient is close to 0, it suggests no correlation between the missing values of the two variables. (d) T-tests were conducted on the missing values to calculate p-values. When the p-value of the T-test is large (typically > 0.05), it can be inferred that the missing data are unrelated to other variables and are missing completely at random. If the T-test shows that the p-values for certain variables are small (typically < 0.05), it indicates a significant difference between the missing data and other variables, suggesting that the missing data are randomly missing but associated with certain features.

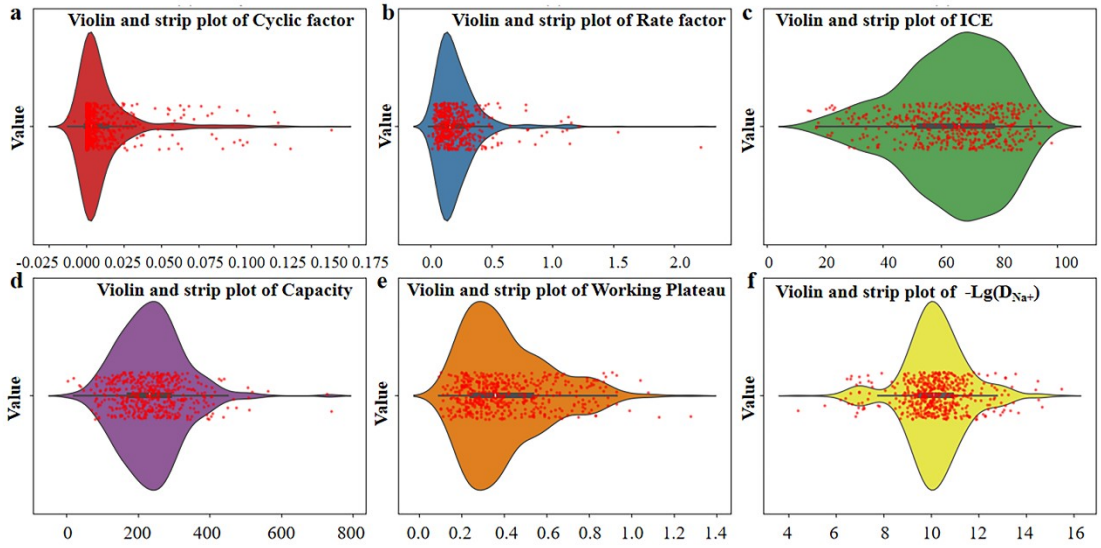

Fig A.2 Data set (filled in) violin plot and dot plot of output features. The horizontal coordinate corresponds to the value of the data point, and the curve reflects the probability density of the data distribution.

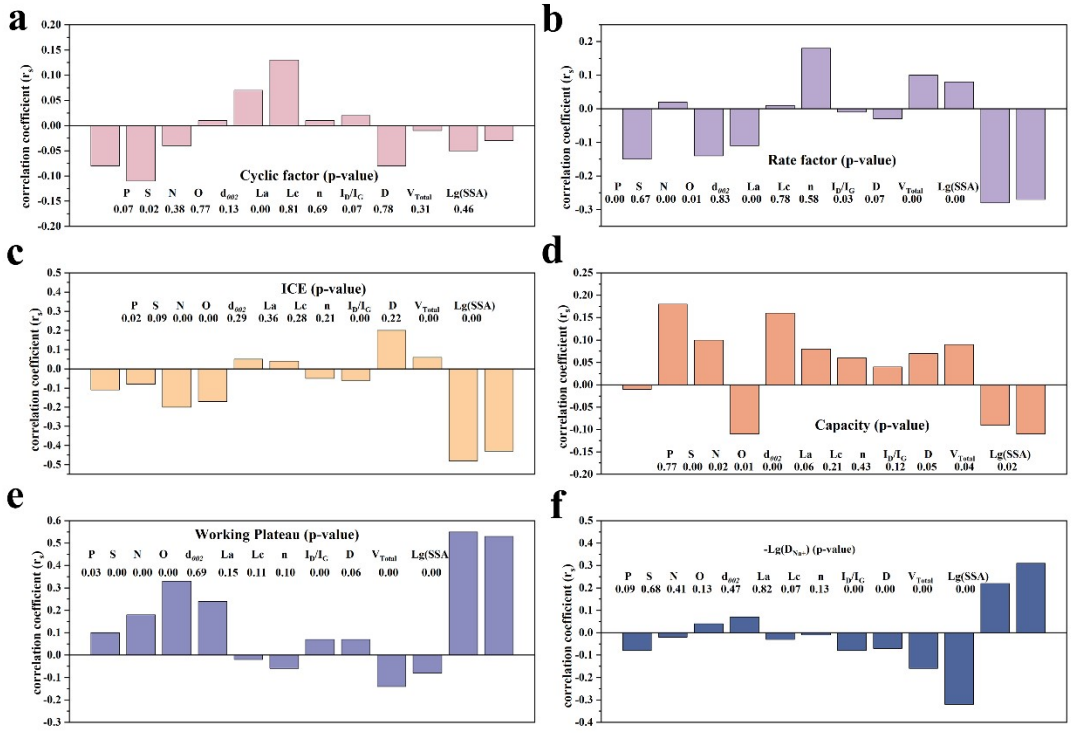

Fig A.3 Spearman rank correlation analysis of structural parameters and discharge performance; The color bars are P, S, N, O, d_{002}、La、Lc、n, I_D/I_G、D, V_{Total}, Lg(SSA) from left to right, and the corresponding significance P-values are inserted in the table.

Table A.5

The fitting results of various regression models for different output targets.

Output target	Model	Train	Val				Test			
		R^2	R^2	RMSE (%)	MSE(%)	MAE(%)	R^2	RMSE (%)	MSE(%)	MAE(%)
Cyclic factor	SVR	0.670	0.244	13.33	1.78	9.82	0.362	10.60	1.12	8.71
	KNN	1.0	0.479	11.06	1.22	8.53	0.292	11.16	1.24	9.11
	XGB	0.972	0.536	10.44	1.09	8.55	0.434	9.97	0.99	8.05
	GBDT	0.999	0.575	9.99	1.00	8.07	0.344	10.74	1.15	8.41
	RF	0.883	0.501	10.83	1.17	8.82	0.527	9.12	0.83	7.10
	Bagging	0.954	0.488	10.97	1.20	8.92	0.382	10.43	1.09	7.46
	CatBoost	0.999	0.707	8.29	0.69	7.22	0.670	7.62	0.58	6.39
	Stacking	0.859	0.637	9.24	0.85	7.87	0.518	9.20	0.85	7.20
	AdaBoost	0.999	0.731	7.96	0.63	5.79	0.643	7.93	0.63	5.87
	ETR	0.995	0.554	10.24	1.05	8.62	0.446	9.88	0.98	7.83
Rate factor	SVR	0.752	0.193	14.64	2.14	1.01	0.107	6.31	0.40	4.88
	KNN	1.0	0.669	9.38	0.88	6.66	0.231	5.86	0.34	4.83
	XGB	0.734	0.365	12.98	1.68	8.85	0.343	5.41	0.29	4.26
	GBDT	0.999	0.680	9.21	0.85	7.12	0.239	5.82	0.34	4.34
	ElasticNet	0.121	0.128	15.21	2.31	10.06	0.096	6.35	0.40	5.11
	RF	0.867	0.634	9.86	0.97	7.65	0.458	4.92	0.24	3.57
	Bagging	0.963	0.706	8.84	0.78	7.16	0.266	5.72	0.33	3.80
	CatBoost	0.999	0.742	8.27	0.68	6.64	0.557	4.44	0.20	3.45
	MLP	0.958	0.518	11.32	1.28	8.35	0.371	5.30	0.28	3.89
	AdaBoost	0.999	0.780	7.64	0.58	5.90	0.749	3.34	0.11	2.33
ICE	SVR	0.526	0.426	10.12	1.02	8.41	0.541	8.23	0.68	7.09
	KNN	1.0	0.548	8.98	0.81	7.02	0.486	8.72	0.76	7.26
	XGB	0.998	0.722	7.04	0.50	5.89	0.702	6.64	0.44	5.39
	GBDT	0.999	0.783	6.22	0.39	5.16	0.738	6.22	0.39	5.12
	ElasticNet	0.219	0.397	10.38	1.08	8.47	0.460	8.93	0.80	7.72
	RF	0.938	0.744	6.76	0.46	5.95	0.720	6.43	0.41	5.45
	Bagging	0.960	0.760	6.55	0.43	5.79	0.749	6.09	0.37	5.25
	CatBoost	0.999	0.761	6.54	0.43	5.27	0.724	6.39	0.41	5.20
	MLP	0.794	0.368	10.63	1.13	8.42	0.268	10.40	1.08	8.25
	AdaBoost	0.999	0.753	6.65	0.44	5.50	0.751	6.07	0.37	4.75
Capacity	SVR	0.391	0.381	7.39	0.55	5.71	0.125	11.35	1.29	8.41
	KNN	1.0	0.499	6.65	0.44	5.57	0.096	11.53	1.33	8.94
	XGB	0.999	0.624	5.77	0.33	4.68	0.388	9.49	0.90	7.26
	GBDT	0.999	0.510	6.58	0.43	5.46	0.355	9.75	0.95	7.37
	Stacking	0.902	0.795	4.25	0.18	2.98	0.733	6.27	0.39	4.25
	RF	0.877	0.431	6.13	0.50	5.39	0.324	9.97	0.99	7.32
	Bagging	0.947	0.509	6.59	0.43	5.18	0.276	10.32	1.06	7.52
	CatBoost	0.999	0.712	5.04	0.25	4.05	0.583	7.83	0.61	5.84

	ETR	0.988	0.597	5.96	0.39	4.90	0.362	9.69	0.94	7.12
	AdaBoost	0.992	0.685	5.28	0.28	4.37	0.424	9.21	0.85	7.23
Working Plateau	SVR	0.846	0.682	9.34	0.87	6.84	0.521	10.71	1.15	8.58
	KNN	1.0	0.713	8.86	0.79	6.67	0.617	9.57	0.92	7.10
	XGB	0.999	0.791	7.57	0.57	5.74	0.634	7.68	0.88	9.36
	GBDT	0.999	0.790	7.58	0.58	6.24	0.847	6.06	0.37	4.72
	ElasticNet	0.497	0.629	10.08	1.02	8.29	0.484	11.12	1.24	8.37
	RF	0.959	0.801	7.38	0.54	6.27	0.798	6.95	0.48	5.36
	Bagging	0.977	0.793	7.54	0.57	6.35	0.812	6.71	0.45	5.18
	CatBoost	0.999	0.785	7.68	0.59	5.80	0.770	7.42	0.55	5.76
	MLP	0.984	0.584	10.68	1.14	8.48	0.456	11.41	1.30	8.30
	AdaBoost	0.999	0.798	7.45	0.56	5.70	0.708	8.36	0.70	6.03
$-LgD_{Na+}$	SVR	0.553	0.527	7.06	0.50	5.45	0.518	7.16	0.51	5.51
	KNN	1.0	0.542	6.95	0.48	5.37	0.265	8.84	0.78	6.82
	XGB	0.997	0.783	4.78	0.23	3.76	0.697	5.68	0.32	4.82
	GBDT	0.999	0.733	5.30	0.28	4.11	0.483	7.41	0.55	6.34
	ElasticNet	0.152	0.454	7.59	0.58	5.36	0.203	9.20	0.85	7.29
	RF	0.891	0.742	5.22	0.27	4.20	0.575	6.72	0.45	5.39
	Bagging	0.960	0.722	5.41	0.29	4.41	0.589	6.61	0.44	5.34
	CatBoost	0.999	0.725	5.39	0.29	4.19	0.635	6.23	0.39	5.16
	MLP	0.702	0.258	8.85	0.78	6.65	0.282	8.73	0.76	6.69
	AdaBoost	0.996	0.746	5.18	0.27	3.91	0.573	6.73	0.45	5.16

Note: Since the Rate and Working Plateau are not normalized, the units of RMSE and MAE calculated by them are ($\times 10^2$h) and ($\times 10^2$v), respectively. The units of MSE are ($\times 10^2$ h^2) and ($\times 10^2$ v^2), respectively.

Table A.6

For the predicted results of initial charge and discharge efficiency (ICE), Capacity (Capacity), and Rate (Rate), the determination coefficient R^2 of the machine learning model on the test set was compared between the reported studies and the present study.

Discharge performance	ICE	Capacity	Rate
Liu [10]et al.	0.60(Bagging)	0.61(XGB)	0.61(GBDT)
Ji[11] et al.	0.71(Bagging)	0.80(Bagging)	0.65(Bagging)
This work	0.75(Bagging)	0.73(Stacking)	0.75(AdaBoost)

Table A.7

The hyperparameter data of the optimal model corresponding to the output targets.

Output target	Optimal model	Hyperparameter data
Cyclic factor	CatBoost	Iterations=1872，depth=8，learning_rate=0.0435，

		l2_leaf_reg=3.84，border_count=132，random_state=42
Rate factor	AdaBoost	estimator__max_depth=15，learning_rate=0.485，n_estimators=437，random_state=42
ICE	Bagging	n_estimators=5000，max_samples=1.0, max_features=0.8956，random_state=42
Capacity	Stacking	catboost__iterations=1000，catboost__depth=9，catboost__learning_rate=0.05，rf__n_estimators=500，svr__c=0.001，random_state=42
Working Plateau	GBDT	n_estimators=925，learning_rate=0.011，max_depth=20，subsample=0.706，max_features=None，min_samples_split=2，min_samples_leaf=1，alpha=0.156，random_state=42
$-LgD_{Na+}$	XGBoost	learning_rate=0.01，max_depth=15，n_estimators=5000，min_child_weight=12，subsample=0.6，colsample_bytree=0.8，lambda=0.1，alpha=0.1，gamma=0

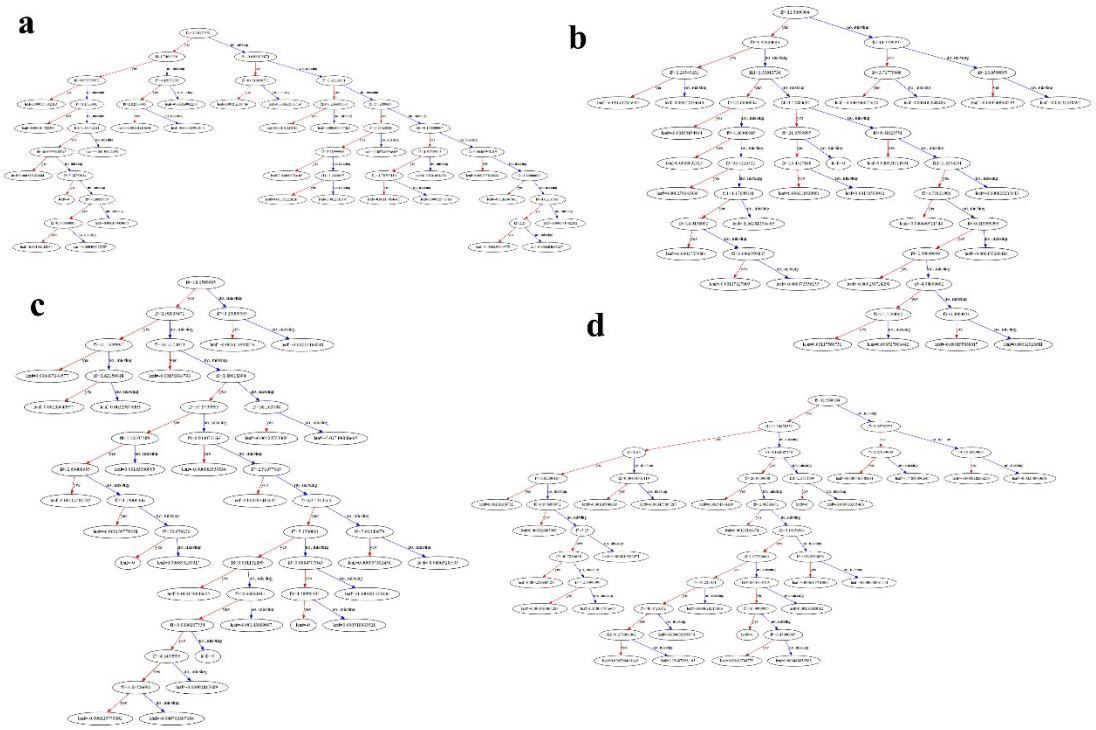

Fig A.4 Diagram of the first 4 decision trees when using the XGB model to predict the relationship between the structure of hard carbon material and D_{Na+} ($-LgD_{Na+}$). f1 to f11 are respectively P, S, N, O, d_{002}、La、Lc、n, I_D/I_G、D, V_{Total}, and Lg(SSA).

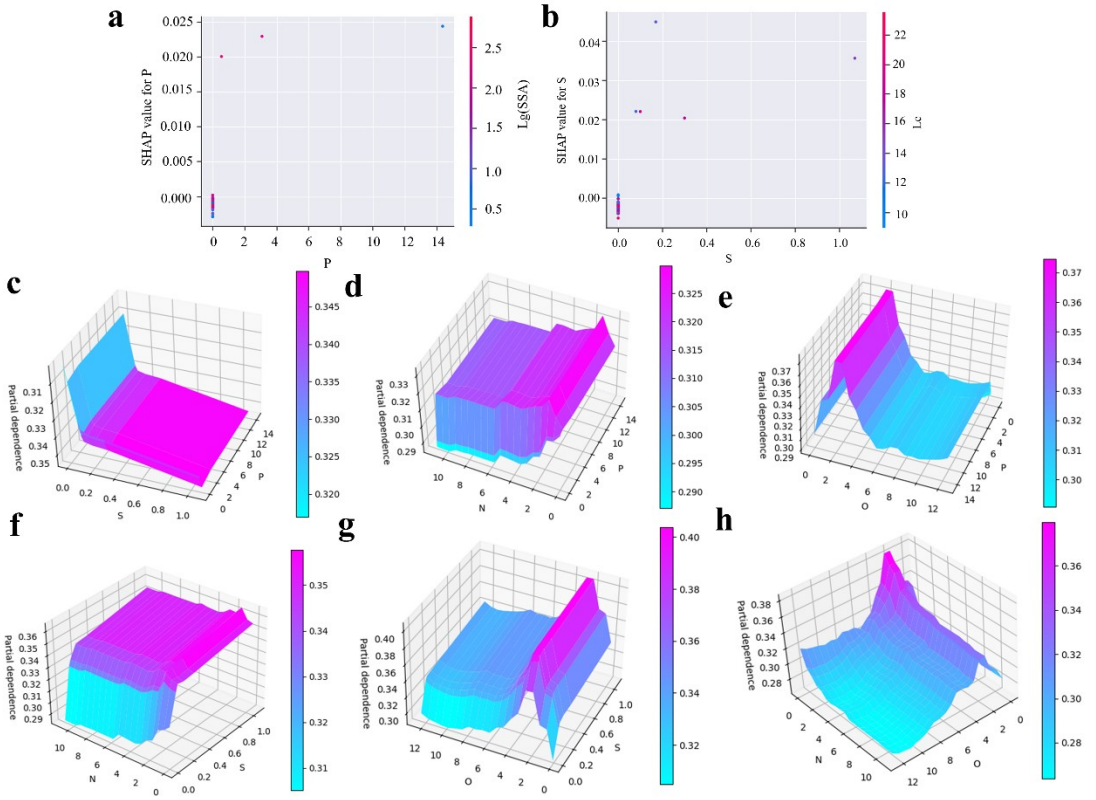

Fig A.5 The impact of heteroatom doping on cycling performance. (a-b) Analysis of the influence of single-atom doping on the cyclic factor based on SHAP values. (a) P, (b) S; the horizontal axis and the right vertical axis correspond to the actual values, while the left vertical axis corresponds to the SHAP values. A SHAP value greater than zero indicates a positive influence on the model's prediction, while a SHAP value less than zero indicates a negative influence. (c-h) The interaction of dual-atom doping on the cyclic factor. (c) P and S; (d) P and N; (e) P and O; (f) S and N; (g) S and O; (h) N and O; the two horizontal axes correspond to the actual values, and the vertical axis corresponds to the magnitude of the interaction.

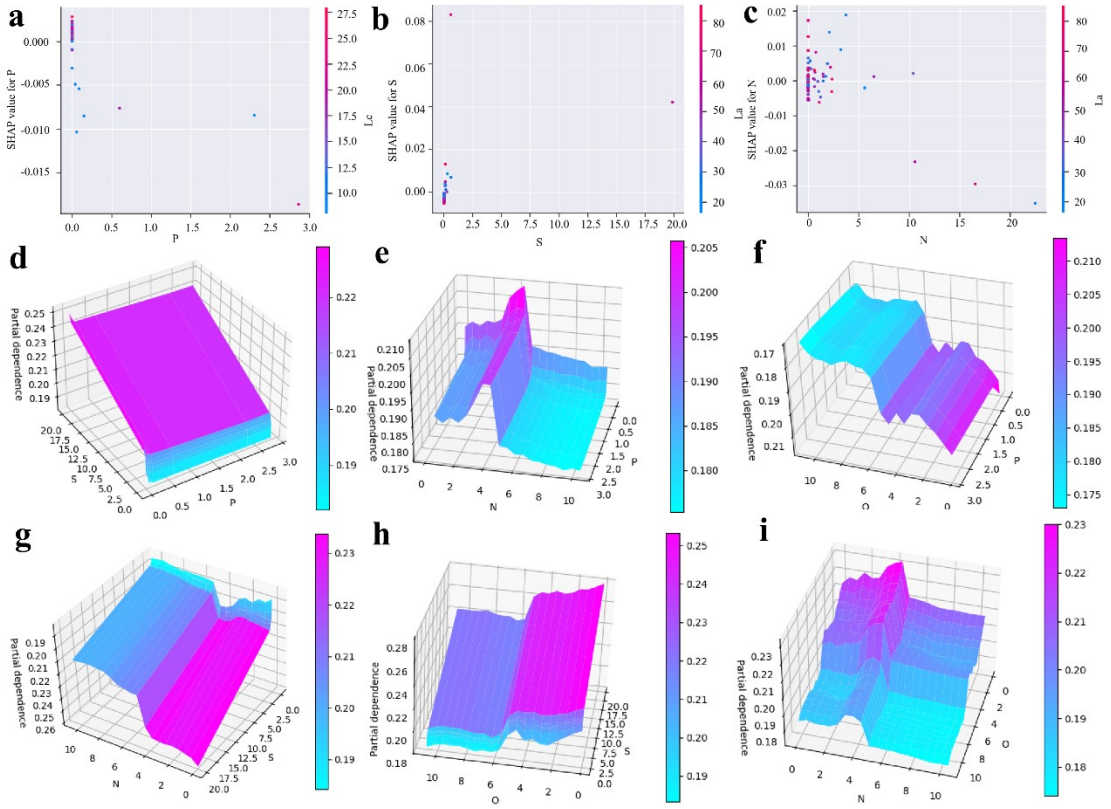

Fig A.6 The impact of heteroatom doping on rate performance. (a-c) Analysis of the influence of single-atom doping on the rate factor based on SHAP values. (a) P, (b) S, (c) N; the horizontal axis and the right vertical axis correspond to the actual values, while the left vertical axis corresponds to the SHAP values. A SHAP value greater than zero indicates a positive influence on the model's prediction, while a SHAP value less than zero indicates a negative influence. (d-i) The interaction of dual-atom doping on the rate factor. (d) P and S; (e) P and N; (f) P and O; (g) S and N; (h) S and O; (i) N and O; the two horizontal axes correspond to the actual values, and the vertical axis corresponds to the magnitude of the interaction.

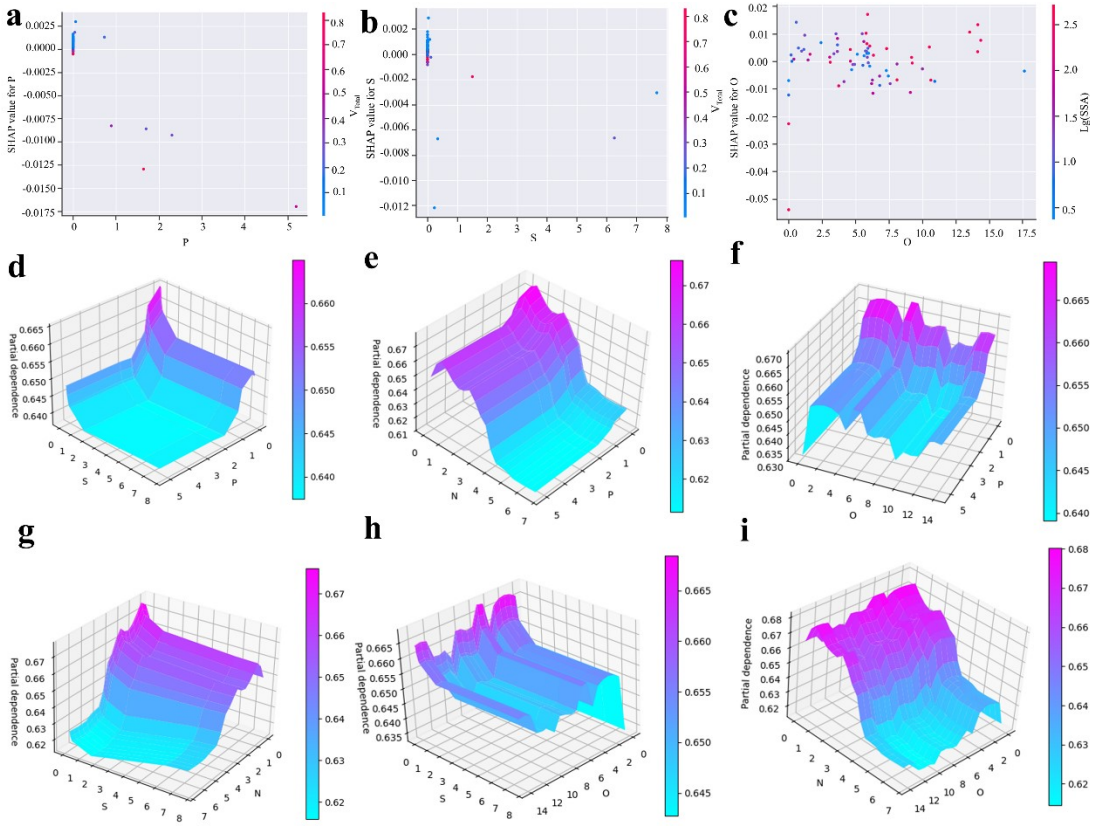

Fig A.7 The impact of heteroatom doping on ICE. (a-c) Analysis of the influence of single-atom doping on the ICE based on SHAP values. (a) P, (b) S, (c) O; the horizontal axis and the right vertical axis correspond to the actual values, while the left vertical axis corresponds to the SHAP values. A SHAP value greater than zero indicates a positive influence on the model's prediction, while a SHAP value less than zero indicates a negative influence. (d-i) The interaction of dual-atom doping on the ICE. (d) P and S; (e) P and N; (f) P and O; (g) S and N; (h) S and O; (i) N and O; the two horizontal axes correspond to the actual values, and the vertical axis corresponds to the magnitude of the interaction.

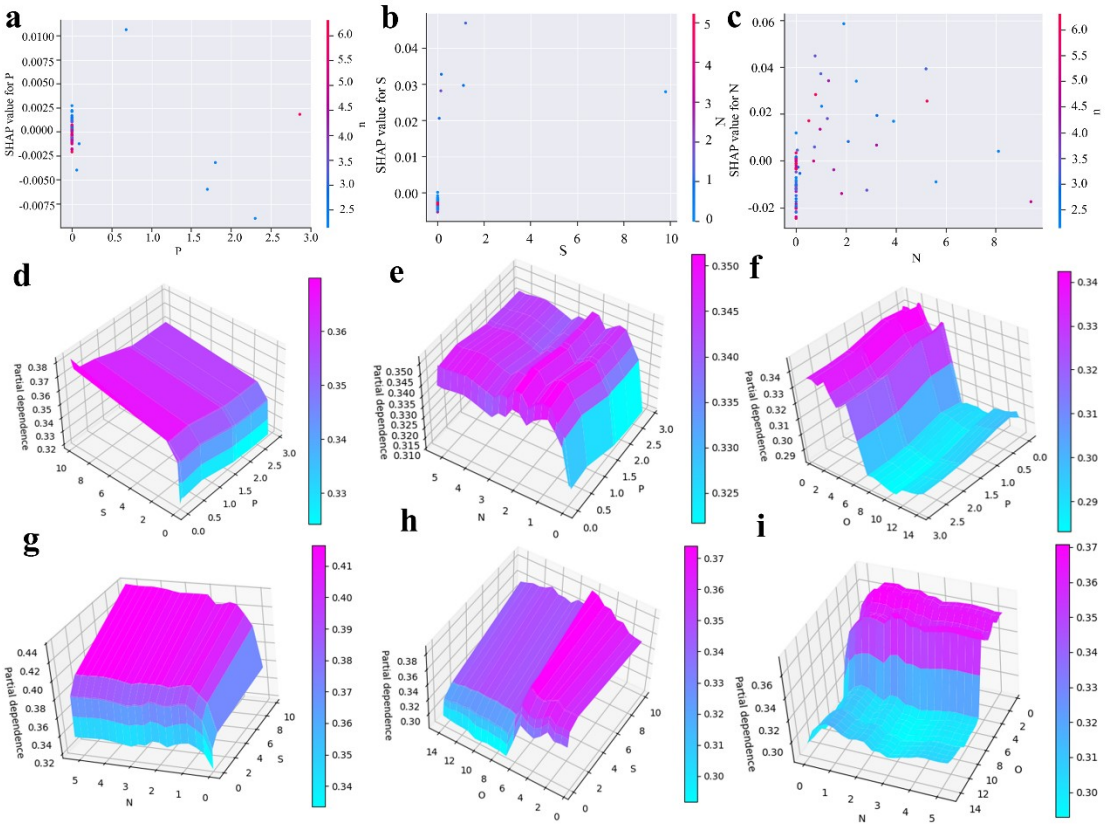

Fig A.8 The impact of heteroatom doping on discharge capacity. (a-c) Analysis of the influence of single-atom doping on the discharge capacity based on SHAP values. (a) P, (b) S, (c) N; the horizontal axis and the right vertical axis correspond to the actual values, while the left vertical axis corresponds to the SHAP values. A SHAP value greater than zero indicates a positive influence on the model's prediction, while a SHAP value less than zero indicates a negative influence. (d-i) The interaction of dual-atom doping on the discharge capacity. (d) P and S; (e) P and N; (f) P and O; (g) S and N; (h) S and O; (i) N and O; the two horizontal axes correspond to the actual values, and the vertical axis corresponds to the magnitude of the interaction.

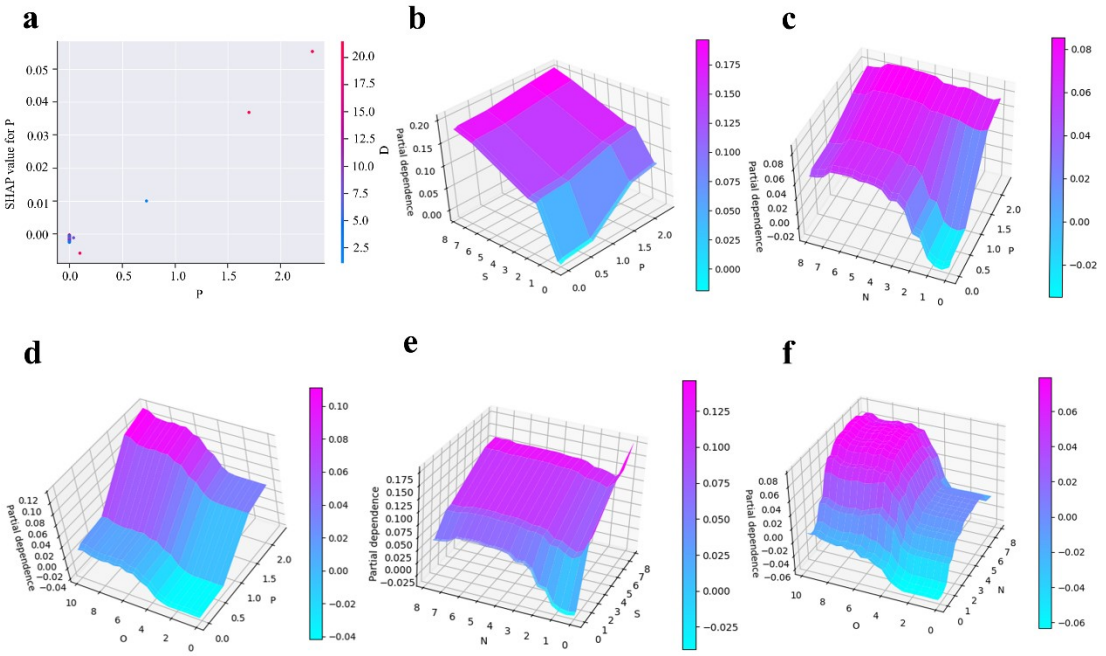

Fig A.9 The impact of heteroatom doping on the working plateau. (a) Analysis of the influence of single-atom doping on the working plateau based on SHAP values. (a) P; the horizontal axis and the right vertical axis correspond to the actual values, while the left vertical axis corresponds to the SHAP values. A SHAP value greater than zero indicates a positive influence on the model's prediction, while a SHAP value less than zero indicates a negative influence. (b-f) The interaction of dual-atom doping on the working plateau. (b) P and S; (c) P and N; (d) P and O; (e) S and N; (f) N and O; the two horizontal axes correspond to the actual values, and the vertical axis corresponds to the magnitude of the interaction.

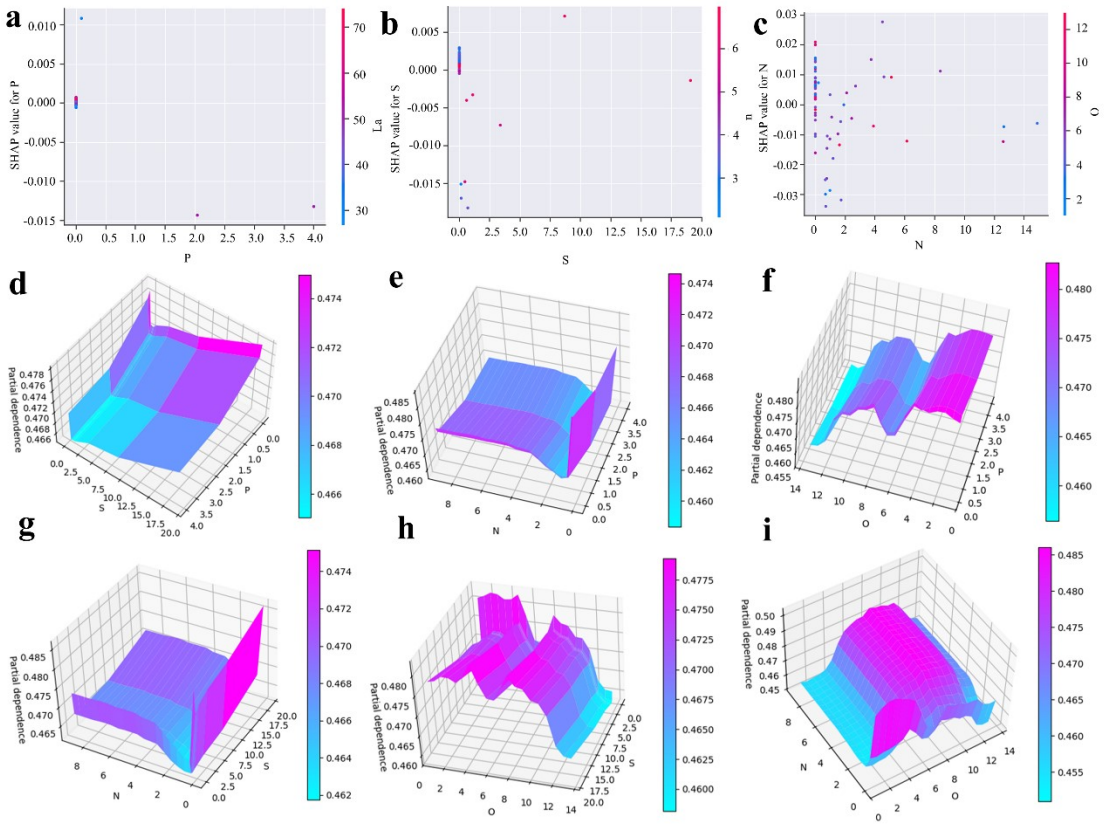

Fig A.10 The impact of heteroatom doping on D_{Na^+}. (a-c) Analysis of the influence of single-atom doping on the $-LgD_{Na^+}$ based on SHAP values. (a) P, (b) S, (c) N; the horizontal axis and the right vertical axis correspond to the actual values, while the left vertical axis corresponds to the SHAP values. A SHAP value greater than zero indicates a positive influence on the model's prediction, while a SHAP value less than zero indicates a negative influence. (d-i) The interaction of dual-atom doping on the $-LgD_{Na^+}$. (d) P and S; (e) P and N; (f) P and O; (g) S and N; (h) S and O; (i) N and O; the two horizontal axes correspond to the actual values, and the vertical axis corresponds to the magnitude of the interaction.

References

1. de Tomas, Carla, Sarat Alabidun, Luke Chater, Matthew T. Darby, Federico Raffone, Paolo Restuccia, Heather Au, Magda M. Titirici, Clotilde S. Cucinotta, and Maria Crespo-Ribadenyra, *Doping carbon electrodes with sulfur achieves reversible sodium ion storage.* Journal of Physics: Energy, 2023. **5**(2).https://doi.org/10.1088/2515-7655/acb570
2. Wang, Hua, SongTao Liu, Cheng Lei, HengRui Qiu, WenQuan Jiang, XueJiao Sun, YongQiang Zhang, and WenXiu He, *P-doped hard carbon material for anode of sodium ion battery was prepared by using polyphosphoric acid modified petroleum asphalt as precursor.* Electrochimica Acta, 2024. **477**.https://doi.org/10.1016/j.electacta.2024.143812
3. Mitchell, Rory and Eibe Frank, *Accelerating the XGBoost algorithm using GPU computing.* PeerJ Computer Science, 2017. **3**.https://doi.org/10.7717/peerj-cs.127
4. Liu, Wanan, Hong Fan, and Meng Xia, *Credit scoring based on tree-enhanced gradient boosting decision trees.* Expert Systems with Applications, 2022. **189**.https://doi.org/10.1016/j.eswa.2021.116034
5. Zilali, Abderrahim, Mehdi Adda, Khaled Ziane, and Maxime Berger, *Machine learning-based state of charge estimation: A comparison between CatBoost model and C-BLSTM-AE model.* Machine Learning with Applications, 2025. **20**.https://doi.org/10.1016/j.mlwa.2025.100629
6. Lin, Shan, Hong Zheng, Bei Han, Yanyan Li, Chao Han, and Wei Li, *Comparative performance of eight ensemble learning approaches for the development of models of slope stability prediction.* Acta Geotechnica, 2022. **17**(4): p. 1477-1502.https://doi.org/10.1007/s11440-021-01440-1
7. Ji, Aibing, Yu Cao, Jinjin Zhang, and Qingqing Li, *SVR-based method for fixed effects interval-valued panel models.* Communications in Statistics - Theory and Methods, 2024. **54**(8): p. 2393-2411.https://doi.org/10.1080/03610926.2024.2369315
8. Arunadevi, M., L. Avinash, Amit Tiwari, C. Durga Prasad, R. Suresh Kumar, L. Raghavendra, G. Veeresha, and Sagarkumar J. Aswar, *Predicting material properties in*

AlSi12Mg alloy additive manufacturing using KNN and ANN machine learning techniques. International Journal on Interactive Design and Manufacturing (IJIDeM), 2025.https://doi.org/10.1007/s12008-025-02276-9

9. Attou, Hanaa, Azidine Guezzaz, Said Benkirane, and Mourade Azrour, *A New Secure Model for Cloud Environments Using RBFNN and AdaBoost.* SN Computer Science, 2025. **6**(2).https://doi.org/10.1007/s42979-025-03691-1
10. Liu, Xiaoxu, Tian Wang, Tianyi Ji, Hui Wang, Hui Liu, Junqi Li, and Dongliang Chao, *Using machine learning to screen non-graphite carbon materials based on Na-ion storage properties.* Journal of Materials Chemistry A, 2022. **10**(14): p. 8031-8046.https://doi.org/10.1039/d1ta10588d
11. Ji, Tianyi, Xiaoxu Liu, Dawei Sheng, Yang Li, Huan Ruan, Hai Guo, Ze Xiang Shen, and Linfei Lai, *Machine learning-assisted thermomechanical coupling fabrication of hard carbon for sodium-ion batteries.* Energy Storage Materials, 2024. **71**.https://doi.org/10.1016/j.ensm.2024.103563