

UC Berkeley

UC Berkeley Previously Published Works

Title

Unveiling the role of lignin in biomass-derived hard carbon anodes via machine learning

Permalink

<https://escholarship.org/uc/item/5fh9m6t1>

Authors

Li, Junxiao

Jin, Yanghao

Sun, Kang

et al.

Publication Date

2025-03-01

DOI

10.1016/j.jpowsour.2025.236323

Peer reviewed

1 Unveiling the Role of Lignin in Biomass-Derived Hard

2 Carbon Anodes via Machine Learning

3 Junxiao Li^a, Yanghao Jin^b, Kang Sun^{a,c*}, Ao Wang^{a,c,d}, Gaoyue Zhang^{a,c}, Limin Zhou^{a,c},

4 Weihong Yang^b, Mengmeng Fan^{a,c}, Jianchun Jiang^{a,c}, Yuming Wen^e, Shule Wang^{f,a,c*}

5

6 ^a Jiangsu Province Key Laboratory of Biomass Energy and Materials, National
7 Engineering Laboratory for Biomass Chemical Utilization, Institute of Chemical
8 Industry of Forest Products, Chinese Academy of Forestry (CAF), Nanjing, 210042,
9 China.

10 ^b Department of Materials Science and Engineering, KTH Royal Institute of
11 Technology, SE-100 44, Stockholm, Sweden.

12 ^c Jiangsu Co-Innovation Center for Efficient Processing and Utilization of Forest
13 Resources, College of Chemical Engineering, Nanjing Forestry University, Longpan
14 Road 159, Nanjing 210037, China.

15 ^d School of Chemistry, Chemical Engineering and Biotechnology, Nanyang
16 Technological University, 62 Nanyang Drive, Singapore, 637459 Singapore.

17 ^e Department of Chemical and Biomolecular Engineering, National University of
18 Singapore, 4 Engineering Drive 4, E5 #02-09, Singapore 117585.

19 ^f Department of Environmental Science. Policy and Management, University of
20 California, Berkeley, CA 94720, USA

21

22 *Email: Kang Sun: W19852858766@163.com; Shule Wang: shule@berkeley.edu.

23 **Abstract**

24 Biomass-derived hard carbon is a sustainable and promising anode material for
25 sodium-ion batteries. Variations in biomass precursors lead to substantial differences in
26 capacity, necessitating a deeper understanding of the underlying mechanisms. This
27 study collected data from 149 relevant literature in the past decade. We used machine
28 learning models to analyze the impact of lignin content and its structure in biomass
29 precursors on the specific capacity of the derived hard carbon. The tree-based ensemble
30 algorithms, particularly XGB and GBDT, showed superior performance, with the
31 optimal model having a R^2 value of up to 0.99 for training and 0.60 for testing.
32 Interpretable machine learning models identified lignin content and its structure as
33 crucial factors, Shapley value analysis highlighted that higher lignin content and well-
34 defined lignin structures positively influence capacity. Also, it is found that optimal
35 pyrolysis temperatures (1000-1400°C) and appropriate retention times are critical for
36 enhancing performance. This work provides insights into optimizing biomass precursor
37 selection and processing for high-performance hard carbon anodes.

38 **Key words:** biomass hard carbon, sodium-ion battery, machine learning, XGB,
39 precursor selection, lignin

40 **1.Introduction**

41 Lithium-ion batteries (LIBs) have been widely commercialized due to their high
42 specific capacity, excellent cycle life, and rate performance [1]. However, the limited
43 availability of lithium resources and their high costs render LIBs less suitable for large-
44 scale energy storage applications [2]. This has driven the exploration of alternative
45 battery technologies that offer similar performance at a lower cost and with more
46 abundant raw materials [3, 4]. Sodium-ion batteries (SIBs) have emerged as one of the
47 most promising alternatives to LIBs due to the low cost and abundant availability of
48 sodium [5]. Despite these advantages, the larger ionic radius of sodium ($1.02\text{\AA}$)
49 compared to lithium ($0.76\text{\AA}$) poses a significant challenge for the intercalation of
50 sodium ions into traditional graphite anodes used in LIBs [6]. This size disparity
51 necessitates the development of novel anode materials to enhance SIB performance [7].

52 Hard carbon, with its low graphitization degree, expansive layer spacing, and well-
53 defined pore structure, is regarded as an ideal anode material for SIBs [8, 9]. Among the
54 various precursors for hard carbon materials, biomass-based sources are particularly
55 attractive due to their wide availability, eco-friendliness, and the extensive research
56 interest they have garnered [10]. Biomass-derived hard carbon anodes have
57 demonstrated significant potential, but the large structural differences among biomass
58 materials and the complexity of their components often result in inconsistent properties
59 in the synthesized hard carbon [11, 12]. This variability underscores the need for a

60 deeper understanding of the relationship between biomass precursor composition and
61 the properties of the resulting hard carbon.

62 Biomass precursors typically consist of cellulose, lignin, and hemicellulose [13].
63 During pyrolysis, the structure and composition of these components critically influence
64 the characteristics of the resulting hard carbon [11]. Studies have shown that the lignin
65 content in biomass plays a crucial role in determining the structural and electrochemical
66 properties of hard carbon [14]. Nita et al. [15] demonstrated that hard carbon derived
67 from walnut shells, which have a high lignin content, exhibited the highest disorder and
68 the best reversible specific capacity among tested samples. The hard carbons prepared
69 from walnut shells showed the highest disorder and the best reversible specific capacity
70 of 315 mAh/g. Dou et al. [16] selected three representative precursors corncobs, peanut
71 shells, and waste apples. Among these the peanut shell with high lignin content yielded
72 hard carbon has the highest specific capacity of 298 mAh/g. Rios et al. [11]
73 systematically studied 25 kinds of biomass precursors, and the results showed that
74 woody and agricultural waste which have high lignin content showed the lowest specific
75 surface area and the highest reversible capacity of 260-314 mAh/g and initial coulombic
76 efficiency (ICE) of 76-87%. In conclusion, the selection of hard carbon precursors is
77 critical in the biomass pyrolysis process and is essential for optimizing this process. A
78 significant factor influencing the properties of biomass precursors is the lignin content
79 and its internal cross-linking structure. Despite these findings, the impact of lignin
80 content and structure on the performance of hard carbon anodes for SIBs remains

81 inadequately explored. Traditional experimental approaches are often limited by the
82 complexity and diversity of biomass materials, as well as the lack of comprehensive
83 data. Machine learning (ML) techniques offer a promising solution to this problem by
84 enabling the analysis of large datasets to identify patterns and predict outcomes based on
85 precursor characteristics [17, 18].

86 Machine learning has been applied to many fields, such as the recognition and
87 classification or calculation of images in medicine [19], the prediction of building loads
88 in architecture [20], and the design of new energy storage materials in materials science
89 [21]. Machine learning has been widely applied in the development of cathode materials
90 for SIBs, expanding our understanding of how ML can be harnessed for different battery
91 components. For example, Kim et al. developed ML models to tackle the structural
92 instability of O3-type layered transition metal oxides (LTMOs) during charge-discharge
93 cycles. By integrating feature engineering techniques with density functional theory
94 (DFT) calculations, they achieved a robust ML classification model with over 95%
95 accuracy in predicting phase stability [22]. Similarly, Oral et al. analyzed a large dataset
96 of 1,227 data points from 335 studies using random forest and decision tree algorithms,
97 successfully predicting key metrics such as discharge capacity and cycle life, and
98 highlighting the importance of electrode type, synthesis method, and structural
99 characteristics in model accuracy [23]. Soundharrajan et al. applied ML to predict
100 crystal stability, which guided DFT validation of structural stability and was followed by
101 experimental synthesis of compounds like NMVTP/C, demonstrating an efficient

pathway for materials discovery [24]. Sun et al. addressed sodium-ion batteries' unique electrochemical behavior through a hybrid deep learning model combining BiLSTM and N-BEATS networks to improve state of charge (SOC) estimation accuracy by adapting time-series predictions to specific electrochemical patterns [25].

In the field of non-graphitic carbon anode of sodium-ion battery biomass, machine learning is also involved. Liu et al. [26] established a structural attribute database based on the crystallinity parameters (L_a , L_c , d_{002} , and I_D/I_G) and main electrochemical properties of non-graphitic carbon, and combined the data analysis results with thermodynamic and kinetic analysis to systematically evaluate the impact of these parameters on the performance of sodium storage. Ji et al. [27] employed machine learning methods to predict the outcomes of 60 experimental groups, excluding data points that did not align with the principles of dynamics and thermodynamics, and the predictions were further validated through experimental methods. Simultaneously, the effects of temperature, pressure, and various structural parameters on the properties of carbon were assessed. Existing research primarily focuses on predicting the relationship between structure and performance, while investigations into raw materials remain lacking. The selection of raw materials can influence the structure of the carbon, subsequently impacting its electrochemical performance. However, there has been no comprehensive analysis of how biomass precursor types, particularly lignin content and structure, influence the performance of hard carbon anodes in SIBs. Given lignin's significant role in determining the properties of biomass-derived hard carbon, it is

crucial to employ ML techniques to screen and optimize precursor selection. This approach can significantly enhance the synthesis process, leading to the production of high-quality hard carbon materials tailored for superior performance in sodium-ion batteries.

In this study, we address the gap in understanding the role of lignin in biomass-derived hard carbon anodes for SIBs by employing ML techniques. Without the need for experimental procedures, it also provides process guidance for the preparation of biomass-derived anode materials for sodium-ion batteries. We collected and analyzed data from 149 studies on biomass-derived hard carbons to develop predictive models that identify key factors influencing hard carbon performance. Our findings offer valuable insights into the selection of biomass precursors and the optimization of process parameters, which can significantly reduce the exploration costs associated with developing sodium-ion anodes and enhance the performance of hard carbon anodes in sodium-ion batteries.

2.Methodology

2.1 Dataset collection

To investigate the role of lignin in biomass-derived hard carbon anodes for sodium-ion batteries (SIBs), we conducted a comprehensive data collection from the Web of Science database using the keywords "biomass," "hard carbon," and "sodium-ion

battery." Our initial search yielded 839 data points from 149 peer-reviewed articles, focusing on studies published within the last decade. To ensure the dataset's reliability and avoid potential bias, we excluded all entries involving activation treatments. Additionally, laboratory experts manually reviewed each entry for relevance and consistency, resulting in a refined dataset of 442 data points. This carefully curated dataset provides a robust foundation for analyzing lignin's influence on the performance of hard carbon anodes in SIBs (shown as Supplementary S2).

Data Inclusion Criteria:

1. Precursor Characteristics: The studies must report detailed information on the type of biomass precursors. We define the lignin content and structural characteristics by examining the precursor's name and description.
2. Process Parameters: The studies must provide comprehensive details on the pyrolysis and pre-treatment processes, including temperatures, holding times, and leaching methods.
3. Structural Parameters: The studies must include data on the structural attributes of the resulting hard carbon, such as interlayer spacing (d_{002}), specific surface area (SSA), pore volume, I_D/I_G ratio, L_a , and L_c .
4. Performance Metrics: The studies must report electrochemical performance metrics, specifically the specific capacity and initial coulombic efficiency (ICE) of the hard carbon anodes.

Data Preprocessing:

1. Standardization: We standardized the units of measurement across different studies to ensure consistency. For example, specific capacity values were converted to mAh/g, and temperature values were converted to degrees Celsius (°C). This unit standardization minimized discrepancies caused by variations in reporting standards across sources. Additionally, we applied StandardScaler from the Scikit-learn library to further standardize the dataset. This scaler transforms each feature to have a mean of zero and a standard deviation of one, which is particularly effective for features with diverse scales, enhancing model convergence and comparability across features.

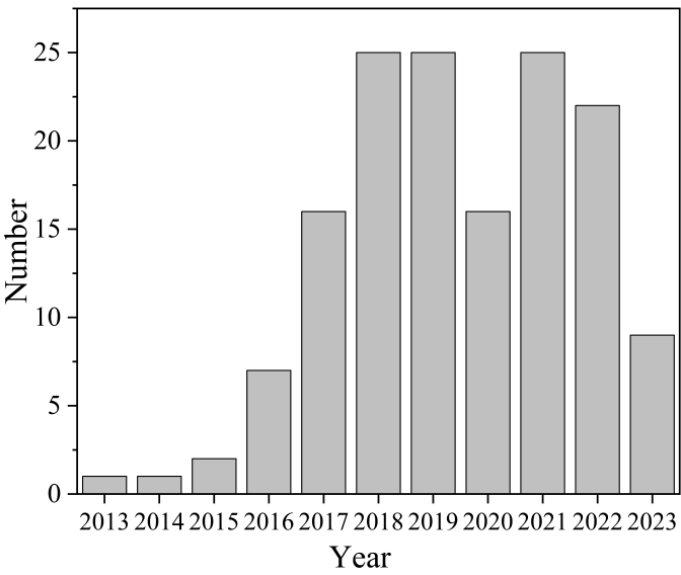
2. Handling Missing Values: For studies with incomplete data, we applied statistical methods such as mean imputation to fill in missing numerical values when the proportion of missing data was manageable. Specifically, features with less than 20% missing data were imputed with mean values to retain as much information as possible. For features with over 20% missing data, we excluded them from the analysis to maintain data integrity. Special attention was given to sensitive data such as doping elements and methods, where mean imputation was applied with caution to minimize potential bias.

3. Categorization: We categorized the collected data into five main groups: precursor parameters (lignin content and structural grade), process parameters (pre-carbonization temperature, peak temperature, holding time, pre-carbonization leaching, post-carbonization leaching, and pre-hydrolysis),

185 structural parameters (d_{002} , SSA, pore volume, I_D/I_G , L_a , L_c), test parameters
186 (current density), and performance parameters (specific capacity).

187 **Data Distribution:**

188 To understand the distribution of our dataset, we analyzed the temporal trends and
189 geographical distribution of the collected studies. Figure 1 shows the number of
190 publications per year, indicating a concentration of relevant studies in the last seven
191 years, with a more even distribution in recent years. This trend highlights the growing
192 interest in biomass-derived hard carbon for SIBs and the increasing application of
193 machine learning in this field.



194
195 Figure 1. Temporal Distribution of Publications on Biomass-Derived Hard Carbon
196 Anodes for Sodium-Ion Batteries (2013-2023)

197

2.2 Feature engineering

In the context of machine learning, feature engineering involves the creation and selection of relevant features (variables) that serve as inputs for the predictive models. For this study, we focused on defining features that capture the characteristics of biomass precursors, process parameters, structural attributes, and performance metrics. The following sections describe the detailed process of creating and defining these features.

Precursor Parameters:

Lignin Content: Lignin content is categorized into six grades based on its proportion in the biomass precursor [28]. The grades are defined as follows:

Grade 0: No lignin (e.g., resin, petroleum-based substances, or pure cellulose)

Grade 1: Grass biomass (lignin content 0-10%)

Grade 2: Leaf biomass (lignin content 5-25%)

Grade 3: Wood biomass (lignin content 20-35%)

Grade 4: Fruit shell biomass (lignin content 20-57%)

Grade 5: Pure lignin

Structural Grade: The structural grade of the biomass is defined by its lignin content and cross-linking complexity:

Grade 0: Non-biomass raw materials (e.g., resins, petroleum-based substances)

Grade 1: Biomass without lignin (e.g., cotton, starch)

Grade 2: Pure lignin with extensive molecular dissociation

Grade 3: Biomass containing lignin with natural cross-linking

219 **Process Parameters:**

220 Pre-carbonization Temperature: The temperature at which the biomass undergoes
221 initial carbonization.

222 Peak Temperature: The maximum temperature reached during the pyrolysis
223 process.

224 Holding Time: The duration for which the biomass is maintained at the peak
225 temperature.

226 Pre-carbonization Leaching: The method used to leach the biomass before
227 carbonization, categorized as:

228 0: No leaching

229 1: Water leaching

230 2: Organic solvent leaching

231 3: Acid/alkali leaching

232 Post-carbonization Leaching: The method used to leach the biomass after
233 carbonization, categorized as:

234 0: No leaching

235 1: Water/organic solvent leaching

236 2: Acid/alkali leaching

237 Pre-hydrolysis: Whether the biomass underwent hydrolysis before carbonization (0:
238 No, 1: Yes)

239 **Structural Parameters:**

240 Interlayer Spacing (d_{002}): The distance between carbon layers in the hard carbon

241 material.

242 Specific Surface Area (SSA): The total surface area per unit mass of hard carbon.

243 Pore Volume: The total volume of pores within the hard carbon.

244 I_D/I_G Ratio: The intensity ratio of the D peak to the G peak in Raman spectra,

245 indicating the degree of disorder.

246 L_a : The average size of carbon crystals along the a-axis.

247 L_c : The thickness of carbon layers stacked along the c-axis.

248 **Test Parameters:**

249 Current Density: The electrical current applied per unit area during electrochemical

250 testing.

251 **Performance Parameters:**

252 Specific Capacity: The amount of charge stored per unit mass of the hard carbon

253 anode.

254 These features were carefully selected and defined to capture the essential

255 characteristics influencing the performance of biomass-derived hard carbon anodes. By

256 standardizing and categorizing these features, we ensured the consistency and reliability

257 of the data used in our machine learning models.

258

259 2.3 Model building and optimization

260 The development of predictive models for assessing the performance of biomass-

261 derived hard carbon anodes involves several critical steps, including the selection of

suitable machine learning algorithms, the training of models on the preprocessed dataset, and the fine-tuning of hyperparameters to enhance model performance. This section outlines the methodologies employed in building and optimizing these models.

Selection of Machine Learning Algorithms: To comprehensively evaluate the factors influencing hard carbon performance, we implemented a diverse array of machine learning algorithms to capture a range of linear and non-linear relationships within the dataset. The selected algorithms include:

- **Linear Models:** Multiple Linear Regression (MLR), Lasso Regression (LASSO), Ridge Regression (RR), and Elastic Net (ENET) are used to identify straightforward relationships between features and hard carbon performance. These models also provide interpretability, allowing us to assess the relative influence of individual features.
- **Tree-Based Models:** Decision Trees (DT), Random Forest (RF), Gradient Boosting Decision Trees (GBDT), and eXtreme Gradient Boosting (XGB) are powerful tools for modeling complex interactions between variables, particularly in datasets with non-linear patterns, which are common in material science applications.
- **Ensemble Methods:** AdaBoost (ADA) and Light Gradient Boosting Machine (LGB) enhance predictive accuracy by combining the outputs of multiple weak learners, thus reducing overfitting and increasing robustness.
- **Others:** K-Nearest Neighbors (KNN) and Support Vector Regression (SVR)

provide additional flexibility for capturing unique data patterns and contribute to the overall robustness of our analysis.

This diverse selection of models ensures a comprehensive exploration of both simple and complex relationships within the dataset, providing a more complete understanding of the factors that influence hard carbon performance in sodium-ion batteries.

Training and Validation: The dataset, comprising 837 data points from 149 studies, was divided into training and validation subsets using a stratified 5-fold cross-validation approach. This method ensures that each fold is representative of the entire dataset, thereby mitigating the risk of overfitting and providing a robust evaluation of model performance. The training process involved the following steps:

1. **Data Normalization:** The StandardScaler from Scikit-learn was applied to normalize the dataset, ensuring that features with different scales do not disproportionately influence the model.
2. **Model Training:** Each selected algorithm was trained on the training subset. The models learned the intricate relationships between the input features (precursor, process, structural, and test parameters) and the output performance metrics (specific capacity and ICE).
3. **Cross-Validation:** To assess and validate the generalizability of the models, we employed 5-fold cross-validation. In this technique, the dataset is divided into five equal subsets. For each fold, 80% of the data is used for training, and the

remaining 20% is used for testing. This process is repeated five times, with each subset serving as the test set once. The final model performance is evaluated by averaging the test scores across all five folds, providing a robust and unbiased measure of the model's predictive accuracy.

Hyperparameter Optimization: Hyperparameter tuning is crucial for enhancing the predictive accuracy and robustness of machine learning models. We employed Bayesian optimization, a method that leverages probabilistic models to efficiently explore the hyperparameter space and identify optimal configurations. Compared to traditional methods such as grid search and random search, Bayesian optimization can achieve optimal results with fewer iterations, as it prioritizes promising regions based on previous evaluations. This approach allows for a more efficient and targeted search, particularly beneficial for models with multiple hyperparameters. The selected hyperparameters are shown in Table S1 in Supplementary S1. The optimization process was carried out as follows:

1. **Initialization:** A Gaussian process was used to model the objective function (model performance) based on the hyperparameters.
2. **Exploration and Exploitation:** The optimization algorithm balanced exploration (searching new hyperparameter values) and exploitation (refining known good values) to iteratively converge on the optimal hyperparameters.
3. **Evaluation:** The performance of each hyperparameter configuration was assessed using the cross-validation score, with the aim of maximizing metrics

such as the coefficient of determination (R^2) and minimizing the mean squared error (MSE).

2.4 Model Interpretation and Software Development

Model Interpretation:

To ensure that our predictive models are not only accurate but also interpretable, we utilized SHapley Additive exPlanations (SHAP) to analyze the contribution of each feature to the model's predictions. SHAP values provide a unified measure of feature importance, revealing how much each feature contributes to a particular prediction [29].

The key advantages of using SHAP include:

Consistency: SHAP values ensure that features contributing more to a model's prediction have higher importance values.

Local and Global Interpretability: SHAP values offer insights at both the global level (overall feature importance across the dataset) and the local level (feature contributions to individual predictions).

The SHAP analysis involves the following steps: 1. Calculation of SHAP Values: For each prediction, the SHAP value of a feature is computed by comparing the model's output with and without the feature. This process quantifies the contribution of each feature to the prediction. 2. Visualization: SHAP summary plots and dependence plots were generated to visualize feature importance and interactions. The summary plot shows the impact of all features across all predictions, while dependence plots illustrate

the effect of a single feature on the prediction. 3. By leveraging SHAP values, we gained a comprehensive understanding of how different features influence the performance of biomass-derived hard carbon anodes, thereby facilitating more informed decision-making in precursor selection and process optimization.

Software Development:

To make the predictive models accessible and user-friendly, we developed a standalone software application using PyQt5 and PyInstaller. This application allows users to input the features of biomass precursors and processing parameters to predict the specific capacity of hard carbon anodes. The development process involved the following steps: 1. User Interface Design: Using PyQt5, we created a graphical user interface (GUI) that is intuitive and easy to navigate. The interface includes fields for inputting precursor parameters (e.g., lignin content, structural grade), process parameters (e.g., pre-carbonization temperature, peak temperature), and structural parameters (e.g., d002, SSA). 2. Model Integration: The trained machine learning models were integrated into the software. The models were serialized using the Joblib library, ensuring efficient loading and execution within the application. 3. Prediction and Output: The application takes user inputs, processes them through the integrated models, and displays the predicted specific capacity. The output includes a detailed breakdown of feature contributions based on SHAP values, providing transparency and interpretability for each prediction. 4. Packaging and Distribution: The application was packaged into a standalone executable using PyInstaller, allowing it to be distributed and

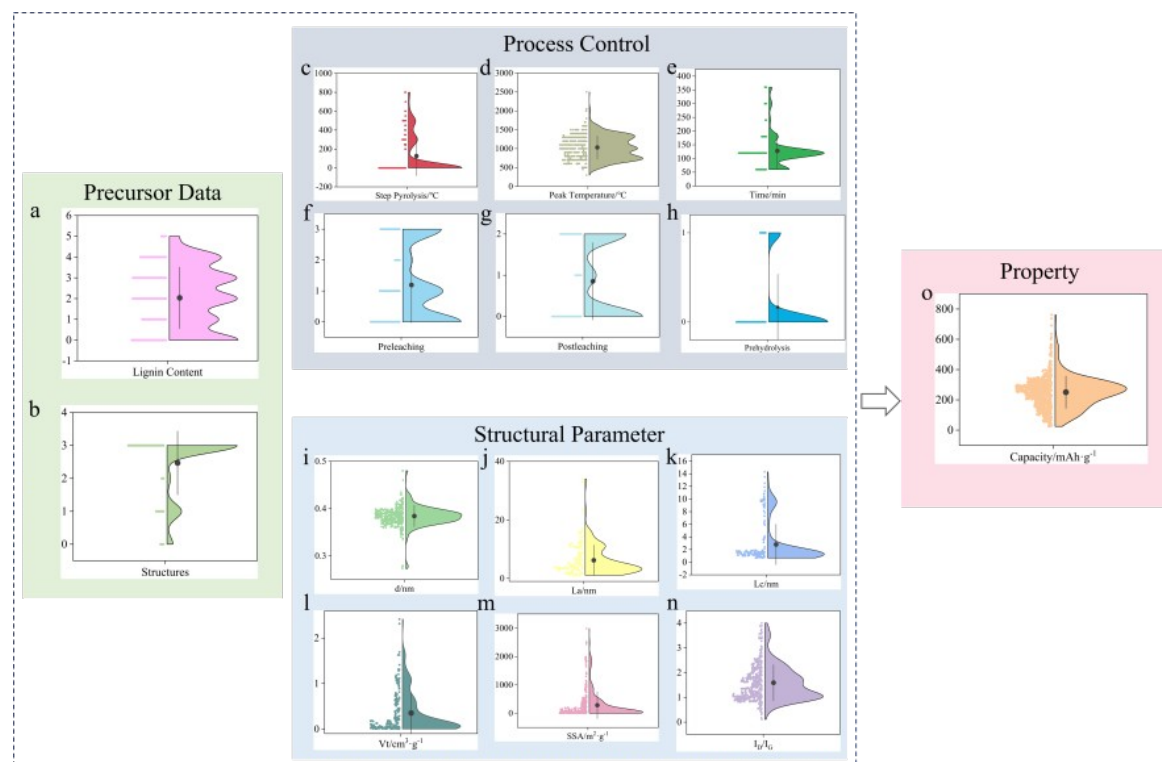
367 run on various operating systems without requiring a Python environment.

368 The software serves as a practical tool for researchers and industry professionals,
369 enabling them to leverage advanced machine learning models to predict the performance
370 of biomass-derived hard carbon anodes. This facilitates the rapid screening of biomass
371 precursors and optimization of processing parameters, ultimately contributing to the
372 development of high-performance sodium-ion batteries.

373

374 **3. Result and discussion**

375 **3.1 Dataset analysis**



376

377 Figure 2. Comprehensive Violin Plot Analysis of Key Parameters in Biomass-Derived Hard Carbon Studies. a) Lignin Content. b) Structural
 378 Grades. c) Step Pyrolysis Temperature. d) Peak Temperature. e) Holding Time. f) Pre-carbonization Leaching Methods. g) Post-carbonization
 379 Leaching Methods. h) Pre-hydrolysis. i) Interlayer Spacing (d₀₀₂). j) Average Carbon Crystal Size (L_a). k) Thickness of Carbon Layers (L_c). l) Total

380 Pore Volume (Vt). m) Specific Surface Area (SSA). n) ID/IG Ratio. o) Specific Capacity (mAh/g).

A total of 837 datasets were gathered from 149 articles, with parameters depicted in half-violin plots to illustrate their magnitude and dispersion (Figure 2a-o). Parameters a-n represent input variables, while o represents the output variable, specific capacity. Input parameters include precursor data (a-b), process control (c-h), and structural parameters (i-n), all of which influence the performance of biomass-derived hard carbon anodes for sodium-ion batteries.

Precursor Data:

Figure 2a illustrates the distribution of lignin content in raw materials, predominantly within grades 0-4, with fewer data points in grade 5. This uniform distribution within grades 0-4 suggests a balanced selection of raw materials. Few studies have chosen pure lignin as raw material because previous studies have shown that the direct pyrolysis of pure lignin to produce hard carbon tends to create long-range ordered structures, which are detrimental to sodium ion storage.[30] Figure 2b shows the structural grade distribution, with grade 3 (biomass containing naturally cross-linked lignin) being the most common, indicating a preference for naturally structured biomass. Biomass sources are abundant and renewable, making them valuable high-quality carbon sources found in nature.

Process Control:

Figure 2c shows step pyrolysis temperatures ranging from 200°C to 800°C, with common peaks at 300°C and 500°C. Adamson et al. found that too high the pre-carbonization temperature would lead to too high the graphitization degree of hard carbon and reduce the layer spacing, which is not conducive to sodium ion storage.

Controlling the temperature between 300°C and 500°C is conducive to building a reasonable hard carbon structure.[31] Therefore, in a large number of experiments, 300-500°C is the preferred step pyrolysis temperature, which is consistent with our conclusion. Figure 2d illustrates peak temperatures, predominantly around 1000°C, a typical range for optimizing hard carbon properties. Figure 2e shows holding times, with 120 minutes being the most common, indicating a standard duration for peak pyrolysis. Figures 2f and 2g display the distribution of pre- and post-leaching methods, with water and organic solvents frequently used. Figure 2h shows that pre-hydrolysis is less commonly applied, indicating its selective use based on specific precursor properties.

Structural Parameters:

Figures 2i-n depict various structural parameters. Figure 2i shows a dense clustering of d_{002} values around 0.385 nm, suggesting an optimal spacing for sodium storage. Figures 2j and 2k indicate that L_a (average size of carbon crystals along the a-axis) and L_c (thickness of carbon layers) are mostly within 0-5 nm and 0-2 nm, respectively. Figure 2l shows pore volume predominantly between 0-0.2 cm³/g, with some scattered data points. Figure 2m illustrates the SSA distribution, with most values near 0 m²/g but extending up to 3000 m²/g. This indicates that a smaller specific surface area is more advantageous for sodium storage. Figure 2n presents the I_D/I_G ratio, mainly between 1 and 3, indicating a desired level of disorder in the carbon structure.

Performance Data:

Figure 2o illustrates the distribution of specific capacity, clustering around 300 mAh/g. Currently, the specific capacity achieved by most studies is as follows: the direct

425 pyrolysis performance of untreated biomass is approximately 250 mAh/g, while the
426 performance of biomass subjected to simple treatments can reach around 300 mAh/g.
427 Furthermore, with precise design optimization, the performance of hard carbon can
428 exceed 300 mAh/g. The highest capacity recorded is nearly 800 mAh/g at 20mA, while
429 some values fall below 300 mAh/g, attributed to high-rate current densities.

430 This comprehensive dataset analysis provides a solid foundation for constructing
431 machine learning models to explore the impact of various input parameters on the
432 performance of biomass-derived hard carbon anodes. This facilitates the informed
433 selection of appropriate hard carbon precursors and the optimization of processing
434 parameters.

435

3.2 Correlation Analysis

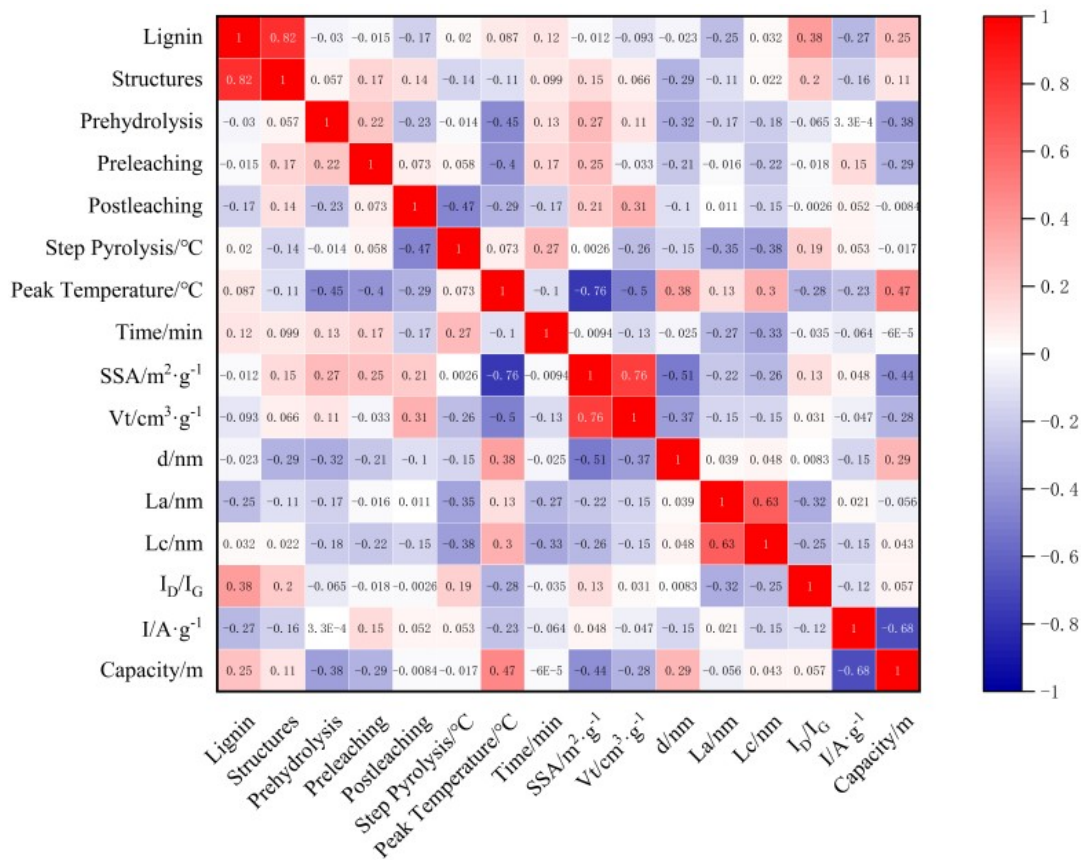


Figure 3. Pearson Correlation Coefficient Heatmap for Key Parameters in Biomass-Derived Hard Carbon Studies

Figure 3 presents the Pearson correlation coefficient heatmap for the collected parameters. The Pearson correlation coefficient measures the linear relationship between two variables, ranging from -1 to 1, where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no correlation [32].

The heatmap reveals several significant correlations:

Lignin Content and Structure: A strong positive correlation ($r=0.82$) is observed between lignin content and structure. This is expected since lignin content is a key factor in defining the structural grade of the biomass precursor. Both lignin content and

structure also show a positive correlation with specific capacity, suggesting that higher lignin content and more structured precursors tend to enhance capacity.

Pre-hydrolysis and Capacity: There is a negative correlation ($r=-0.38$) between pre-hydrolysis and capacity. Pre-hydrolysis degrades cellulose and hemicellulose [33], leading to a lower structural grade and subsequently reducing the capacity of the hard carbon.

Peak Temperature and Capacity: A strong positive correlation ($r=0.47$) exists between peak temperature and capacity. Higher peak temperatures promote the rearrangement of carbon microcrystals and shrinkage of pore structures [34], leading to a favorable "card-house" model for sodium storage, which enhances the capacity.

Testing Current (I) and Capacity: The testing current shows a strong negative correlation with specific capacity ($r=-0.68$). Higher current densities result in less effective internal kinetic reactions within the battery, reducing the specific capacity during testing.

Additionally, the heatmap provides insights into the relationships between process parameters and structural attributes. For instance, it is noted that:

Pre-leaching and Post-leaching: Pre-leaching has a negative impact on capacity, while post-leaching does not exhibit a significant correlation with capacity. This suggests that the timing and method of leaching are critical factors affecting the final performance of the hard carbon.

Step Pyrolysis Temperature and Structural Parameters: Step pyrolysis temperature shows varying effects on different structural parameters. For example,

470 higher temperatures generally reduce specific surface area (SSA) and interlayer spacing
471 (d_{002}), while increasing the amount of closed pores, which collectively influence the
472 capacity [35].

473 While Pearson correlation effectively captures linear relationships, it may not fully
474 elucidate more complex, non-linear interactions between variables. For example, the
475 correlation between ID/IG ratio and capacity is relatively weak ($r=0.057$), despite
476 evidence from previous studies by Qin et al. [36] suggesting that increased I_D/I_G ratios,
477 indicating more defect sites, can enhance sodium storage capacity. This discrepancy
478 highlights the limitations of Pearson correlation in capturing non-linear effects and the
479 need for advanced machine learning models to uncover these intricate relationships.

480 Further research utilizing machine learning regression models is essential to delve
481 deeper into these non-linear associations, providing a more comprehensive
482 understanding of how various factors interact to influence the performance of biomass-
483 derived hard carbon anodes.

484

485

3.3 Model training and hyperparameters tuning

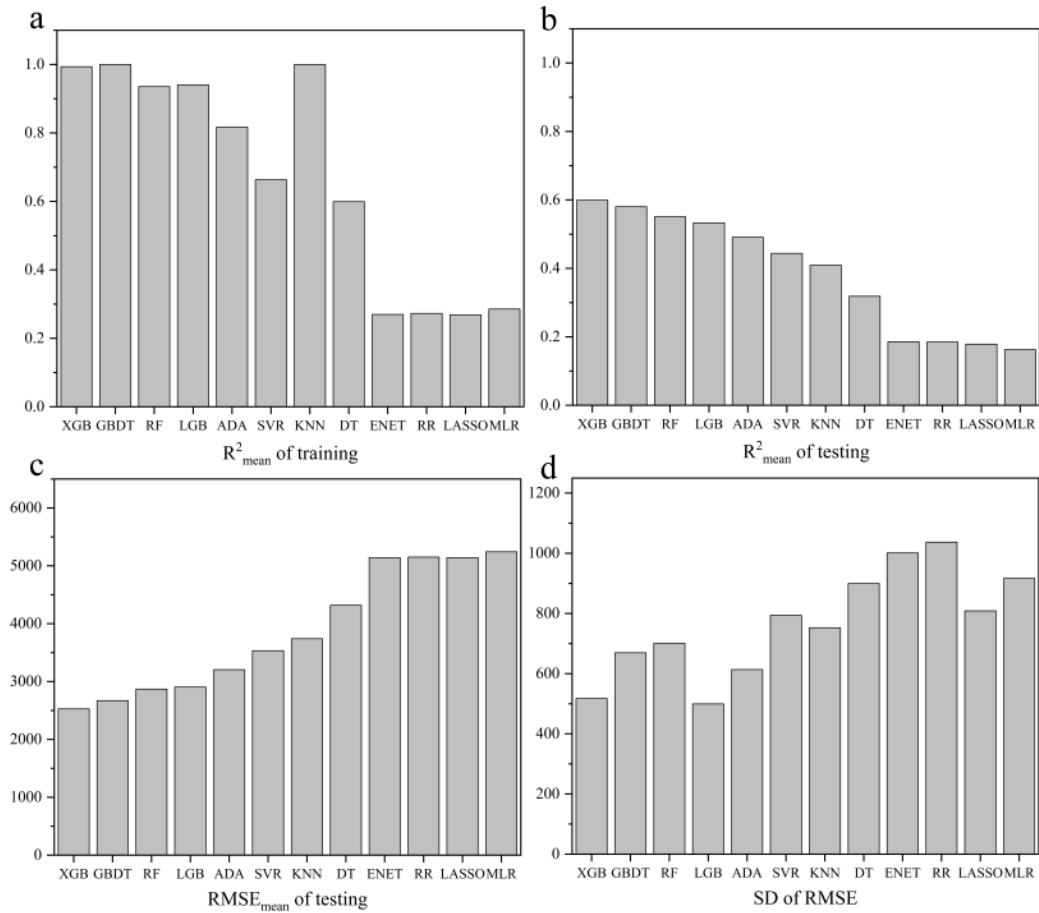


Figure 4. Model Performance Comparison. a) R^2_{mean} of training. b) R^2_{mean} of testing. c) $\text{RMSE}_{\text{mean}}$ of testing. d) Standard Deviation of RMSE.

The model training and hyperparameter tuning were conducted to identify the most effective machine learning algorithms for predicting the performance of biomass-derived hard carbon anodes. Figure 4 illustrates the comparative performance of the models in terms of training and testing R^2 scores, as well as the RMSE (Root Mean Squared Error) and its standard deviation (SD).

Training Results: Figure 4a shows the R^2 scores for the training phase. The eXtreme Gradient Boosting (XGB), Gradient Boosting Decision Trees (GBDT), and

Random Forest (RF) models achieved the highest R^2 scores, indicating an excellent fit with the training data. XGB and GBDT both demonstrated near-perfect R^2 values, reflecting their strong ability to capture complex relationships in the data.

Testing Results: The performance of the testing data, as shown in Figure 4b, provides a clearer picture of model generalizability. XGB and GBDT again emerged as the top performers, with R^2 scores significantly higher than other models, indicating robust predictive capabilities. However, all models experienced a drop in R^2 scores when moving from training to testing data, highlighting the challenge of overfitting.

RMSE Analysis: Figure 4c depicts the RMSE for the testing phase, where lower values indicate better predictive accuracy. XGB and GBDT had the lowest RMSE, further validating their superior performance. Figure 4d shows the standard deviation of RMSE, with XGB and GBDT displaying relatively lower variability, suggesting more consistent predictions across different folds of the cross-validation.

Hyperparameter Tuning: The hyperparameters for each model were fine-tuned using Bayesian optimization, and the optimal combinations are detailed in Supplementary S3. This tuning process was critical for enhancing model performance, particularly for complex algorithms like XGB and GBDT, which benefited from adjustments in learning rate, tree depth, and regularization parameters.

In summary, the results indicate that tree-based ensemble methods, particularly XGB and GBDT, are the most effective for predicting the specific capacity of biomass-derived hard carbon anodes. These models not only achieved high R^2 scores and low RMSE in the testing phase but also demonstrated consistent performance across

different cross-validation folds. This comprehensive evaluation underscores the importance of selecting and fine-tuning machine learning models to achieve reliable and accurate predictions in material science applications.

The superior performance of tree-based models on this dataset can be attributed to several inherent advantages of the algorithms. Firstly, tree-based models such as XGB and GBDT are excellent at capturing non-linear relationships between features and the target variable, which is crucial given the complex interactions present in the dataset. These models are capable of automatically handling feature interactions without the need for explicit specification, making them particularly powerful for datasets with high dimensionality and intricate dependencies. Additionally, tree-based models are robust to outliers and irrelevant features, as they can effectively split the data to minimize impurity, thereby focusing on the most informative features. This characteristic is especially beneficial in the context of material science, where the data can be noisy and contain measurements with varying degrees of precision. Furthermore, ensemble methods like XGB and GBDT combine the predictions of multiple weak learners to form a strong learner, significantly improving the model's accuracy and generalization capability. By iteratively correcting the errors of the previous trees, these algorithms can achieve a high level of predictive performance and reduce the risk of overfitting. Moreover, the use of regularization techniques in these models helps to prevent overfitting by penalizing overly complex models, thus ensuring a balance between bias and variance. The ability to tune multiple hyperparameters, such as learning rate, tree depth, and the number of trees, allows for fine-grained control over the model

complexity and performance, making these algorithms highly adaptable to various types of data.

3.4 Model interpretation

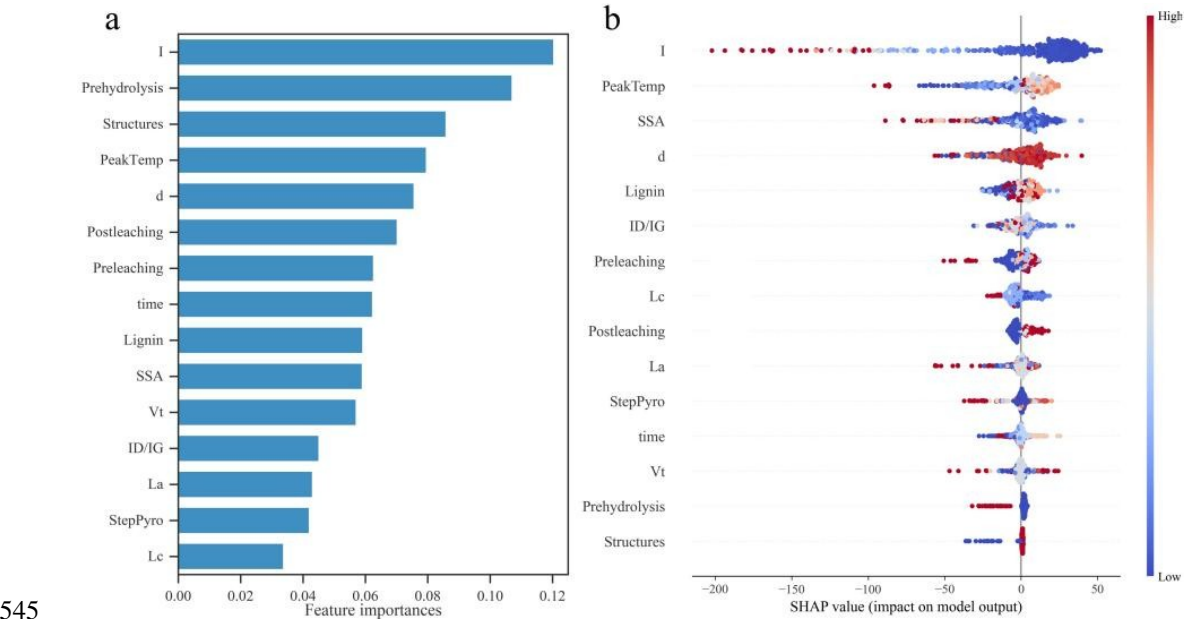


Figure 5. Feature Importance and SHAP Value Analysis for XGB Model. a) Feature importance based on XGB model output. b) SHAP value plot showing the impact of each feature on model output, with red indicating higher feature values and blue indicating lower feature values.

Figure 5 presents the feature importance (a) and SHAP value (b) output by the optimized XGB model. Feature importance reflects the contribution of each feature to the prediction results, while SHAP values provide a detailed interpretation of how each feature impacts individual predictions.

The top three features in terms of feature importance are testing current (I),

556 prehydrolysis, and structures, with importance values of 0.12, 0.11, and 0.086,
557 respectively. Among these, testing current (I) has the most direct and significant impact
558 on capacity. The SHAP values show a strict linear relationship between testing current
559 and capacity: as the current increases, the negative impact on capacity also increases.
560 This serves as a validation reference to ensure that our conclusions are consistent with
561 reality. The high importance of testing current is due to its direct influence on the
562 internal kinetic reactions within the battery, where higher currents reduce the
563 effectiveness of sodium ion intercalation and deintercalation, leading to lower capacity.

564 Both prehydrolysis and structure reflect the selection of the precursor. A higher
565 structural grade positively indicates the required structural characteristics of the
566 precursor, while the negative impact of prehydrolysis underscores the importance of
567 preserving the original dense biomass structure. The SHAP value plot reveals that
568 higher structural levels positively impact capacity, whereas lower structural levels have
569 varying degrees of negative effects on capacity. A low prehydrolysis value, indicating no
570 prehydrolysis of the raw material, positively influences capacity. Conversely, higher
571 prehydrolysis values, indicating prehydrolysis of the raw material, have varying degrees
572 of negative impact on capacity. Lignin content is another critical parameter, as shown by
573 both feature importance and SHAP values. Slightly higher lignin content positively
574 affects capacity because lignin contributes to the structural rigidity and carbon yield
575 during pyrolysis. However, excessively high or low lignin content adversely affects
576 capacity, possibly due to suboptimal structural properties when pure lignin or lignin-
577 deficient biomass is used.

578 The results not only assist in raw material selection but also offer insights into
579 process operations and hard carbon structures. Figure 5(a) shows that the feature
580 importance of pre-leaching and post-leaching impregnation is relatively similar in
581 magnitude, with post-leaching having a slightly greater impact than pre-leaching. The
582 SHAP value results in Figure 5(b) suggest that the effects on capacity are not definitive
583 when employing high-value reagents (such as acids and bases) in pre-leaching. In post-
584 leaching processes, the utilization of acid/base reagents can notably enhance capacity by
585 removing inorganic impurities and increasing microporosity. No leaching will
586 undeniably lead to a decrease in capacity. Previous studies have also compared the types
587 of leaching agents and the sequence of leaching. Tsai et al. [37] compared the effects of
588 non-leaching, water leaching, and acid leaching in post-leaching and found that pore
589 volume increased in samples leached with water and acid. Beda et al. [38] compared the
590 differences in leaching materials before/after pyrolysis and found that leaching after
591 pyrolysis can more effectively remove inorganic alkaline impurities, increase
592 micropores more effectively, and induce contraction in the interlayer spaces of graphite.

593 Another crucial process parameter is the peak temperature. Figure 5(b) suggests
594 that the peak temperature should be maintained within an optimal range; however,
595 excessively high peak temperatures can have a detrimental impact on capacity. This
596 occurs because if the peak temperature is too low, it hinders the reorganization of carbon
597 microcrystals and pore contraction, leading to the formation of porous carbon with
598 increased surface defects [39]. Conversely, excessive peak temperatures result in an
599 overly high graphitization degree of the hard carbon structure, reducing the disorder in

the carbon microcrystalline structure and narrowing the layer spacing, and this scenario is unfavorable for the deintercalation of sodium ions [40, 41]. Based on the range of data we have collected, the optimal pyrolysis temperature should fall between 1000 °C and 1400 °C. In step pyrolysis, it is important to avoid excessively high temperatures as they can adversely affect capacity. Both excessive and insufficient insulation times can have a negative impact on capacity, therefore, it is recommended to select a moderate value within a reasonable range. According to the structural parameter guidelines outlined in Figure 5(b), it is anticipated that the structure of hard carbon should exhibit a smaller SSA, smaller Lc, and larger interlayer spacing. Other process parameters such as La, ID/IG, and Vt do not exhibit a strong unidirectional relationship with capacity within the collected parameter range.

3.5 Software development

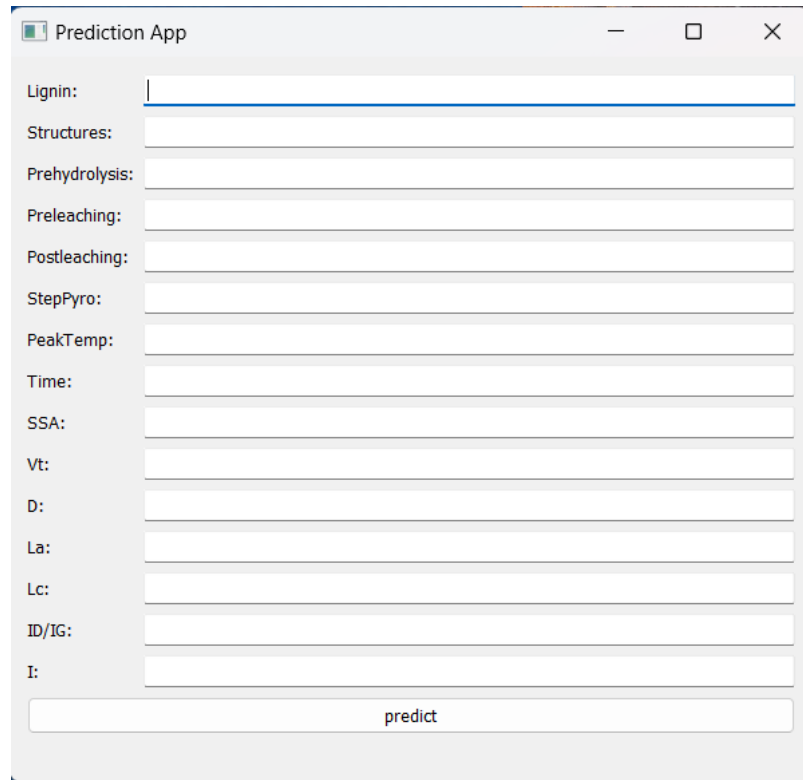


Figure 6. Prediction App Interface. The interface allows users to input features such as lignin content, structural grade, and processing parameters to predict the specific capacity of hard carbon anodes.

To make our predictive models accessible and practical, we developed a user-friendly software application using PyQt5 for the GUI and PyInstaller for packaging. The interface, shown in Figure 6, allows users to input parameters such as lignin content, structural grade, prehydrolysis, preleaching, postleaching, step pyrolysis temperature, peak temperature, holding time, SSA, pore volume (V_t), interlayer spacing (d_{002}), average carbon crystal size (L_a), thickness of carbon layers (L_c), I_D/I_G ratio, and testing current (I).

Users enter these values and click "predict," prompting the application to process

the inputs through the trained XGB model. The predicted specific capacity is then displayed, providing immediate feedback. The code for this software can be found in Supplementary S1.

3.6 Discussion

This study has demonstrated the importance of lignin content and structural level in the selection of biomass precursors for producing high-performance hard carbon materials. Through machine learning analysis and simulation of an extensive dataset, we identified that precursors with higher lignin content and well-defined structures are particularly suitable for hard carbon production. Our focus on biomass-derived precursors highlights the influence of regional availability on raw material selection, underscoring the need for region-specific processing methods. This insight has broad applicability, allowing for tailored approaches to biomass utilization based on local resource characteristics.

Additionally, our findings provide valuable guidance for optimizing process parameters and controlling the structure of hard carbon. Specifically, post-leaching without prehydrolysis, maintaining peak pyrolysis temperatures between 1000 °C and 1400 °C, and implementing moderate retention times are crucial for achieving desirable hard carbon properties. The initial step pyrolysis temperature should also be carefully controlled to avoid excessive temperatures that can adversely affect capacity. The optimal hard carbon structure, as identified in this study, should exhibit a smaller specific surface area (SSA), smaller L_a and L_c values, and larger interlayer spacing

(d002). These recommendations can serve as practical guidelines for future experimental and industrial efforts in developing sodium-ion battery anodes.

3.7 Limitations and Future Research

While the contributions discussed in Section 3.6 mark significant advancements, this study does have inherent limitations. The current model uses precursor, process, and structural parameters to predict performance, simplifying the complex interdependencies present in real-world production processes. Ideally, precursor choice should influence structural characteristics, which in turn dictate performance. Our approach, though effective for initial screening and optimization, does not fully capture this hierarchical relationship. Future research should aim to establish a comprehensive, stepwise pathway that begins with precursor selection, progresses to structural characterization, and ultimately leads to performance prediction.

Additionally, while our machine learning models have demonstrated strong accuracy and robustness, the reliance on existing data could introduce biases or gaps that may impact the results. Experimental validation and a broader range of data sources would further strengthen the models' reliability and applicability. Another limitation is the relatively broad classification of precursors; future studies should incorporate more refined and detailed categorizations of biomass types and properties to improve predictive power.

Addressing these limitations will allow future research to build on our findings, resulting in more nuanced and comprehensive models that better capture the complexities of biomass-derived hard carbon production.

671

672 **4. Conclusion**

673 This study employed machine learning techniques to establish a robust predictive
674 model for the synthesis of high-performance biomass-derived hard carbon anodes for
675 sodium-ion batteries. By incorporating key inputs such as lignin content, structural
676 grade, process parameters, and structural attributes, we utilized 12 different machine
677 learning methods to construct and validate these models. Our analysis, particularly using
678 the XGB model, provided insights into the impact of various parameters on the specific
679 capacity of hard carbon anodes.

680 Key findings from this study include the identification of high lignin content and
681 well-defined structural grades as critical factors in selecting suitable biomass precursors.
682 Optimal process conditions were determined to involve post-leaching with acid or alkali
683 after pyrolysis, maintaining peak temperatures between 1000°C and 1400°C, and
684 moderate retention times to enhance the performance of the resulting hard carbon. The
685 ideal structural attributes were found to be smaller specific surface area (SSA), smaller
686 La and Lc values, and larger interlayer spacing (d_{002}).

687 In addition to guiding raw material selection and process optimization, this study
688 highlighted the limitations of current models in capturing the hierarchical relationships
689 between precursor properties, structural characteristics, and performance. Future
690 research should aim to refine the classification of biomass precursors, develop more
691 detailed predictive frameworks, and incorporate additional experimental validation to

enhance model reliability and applicability.

Overall, this study provides a comprehensive framework for the development of high-performance hard carbon anodes, leveraging machine learning to optimize precursor selection and processing parameters. The findings and methodologies presented herein can significantly advance the field of sodium-ion batteries and promote the efficient utilization of biomass resources.

Acknowledgement

This work was supported by the National Key R&D Program of China (Grant No. 2023YFB4203703). We also gratefully acknowledge the financial support from the Natural Science Foundation of Jiangsu Higher Education Institutions (Grant No. 23KJB220005). The corresponding author, Shule Wang, conducted this research during his tenure at Nanjing Forestry University, prior to his current position at the University of California, Berkeley.

During the preparation of this work the authors used ChatGPT in order to improve the readability and language of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Reference:

- [1] M. Li, J. Lu, Z. Chen, K. Amine, 30 Years of Lithium-Ion Batteries, *Adv Mater*, 30 (2018) e1800561. 10.1002/adma.201800561.
- [2] T.M. Gür, Review of electrical energy storage technologies, materials and systems: challenges and prospects for large-scale grid storage, *Energy & Environmental Science*, 11 (2018) 2696-2767. 10.1039/c8ee01419a.
- [3] F. Wei, Q. Zhang, P. Zhang, W. Tian, K. Dai, L. Zhang, J. Mao, G. Shao, Review—Research Progress on Layered Transition Metal Oxide Cathode Materials for Sodium Ion Batteries, *Journal of The Electrochemical Society*, 168 (2021). 10.1149/1945-7111/abf9bf.
- [4] A. Ramesh, A. Tripathi, P. Balaya, A mini review on cathode materials for sodium-ion batteries, *International Journal of Applied Ceramic Technology*, 19 (2021) 913-923. 10.1111/ijac.13920.
- [5] H. Hou, X. Qiu, W. Wei, Y. Zhang, X. Ji, Carbon Anode Materials for Advanced Sodium-Ion Batteries, *Advanced Energy Materials*, 7 (2017). 10.1002/aenm.201602898.
- [6] S. Qiao, Q. Zhou, M. Ma, H.K. Liu, S.X. Dou, S. Chong, Advanced Anode Materials for Rechargeable Sodium-Ion Batteries, *ACS Nano*, 17 (2023) 11220-11252. 10.1021/acsnano.3c02892.
- [7] E. Irisarri, A. Ponrouch, M.R. Palacin, Review—Hard Carbon Negative Electrode Materials for Sodium-Ion Batteries, *Journal of The Electrochemical Society*, 162 (2015) A2476-A2482. 10.1149/2.0091514jes.

733 [8] M. Wahid, D. Puthusseri, Y. Gawli, N. Sharma, S. Ogale, Hard Carbons for Sodium-
734 Ion Battery Anodes: Synthetic Strategies, Material Properties, and Storage
735 Mechanisms, *ChemSusChem*, 11 (2018) 506-526. 10.1002/cssc.201701664.

736 [9] D. Saurel, B. Orayech, B. Xiao, D. Carriazo, X. Li, T. Rojo, From Charge Storage
737 Mechanism to Performance: A Roadmap toward High Specific Energy Sodium-Ion
738 Batteries through Carbon Anode Optimization, *Advanced Energy Materials*, 8 (2018).
739 10.1002/aenm.201703268.

740 [10] M. Thompson, Q. Xia, Z. Hu, X.S. Zhao, A review on biomass-derived hard carbon
741 materials for sodium-ion batteries, *Materials Advances*, 2 (2021) 5881-5905.
742 10.1039/d1ma00315a.

743 [11] C. del Mar Saavedra Rios, L. Simonin, C.M. Ghimbeu, C. Vaultot, D. da Silva Perez,
744 C. Dupont, Impact of the biomass precursor composition in the hard carbon properties
745 and performance for application in a Na-ion battery, *Fuel Processing Technology*, 231
746 (2022). 10.1016/j.fuproc.2022.107223.

747 [12] X. Dou, I. Hasa, D. Saurel, C. Vaalma, L. Wu, D. Buchholz, D. Bresser, S. Komaba,
748 S. Passerini, Hard carbons for sodium-ion batteries: Structure, analysis, sustainability,
749 and electrochemistry, *Materials Today*, 23 (2019) 87-104.
750 10.1016/j.mattod.2018.12.040.

751 [13] F.-X. Collard, J. Blin, A review on pyrolysis of biomass constituents: Mechanisms
752 and composition of the products obtained from the conversion of cellulose,
753 hemicelluloses and lignin, *Renewable and Sustainable Energy Reviews*, 38 (2014) 594-
754 608. 10.1016/j.rser.2014.06.013.

755 [14] C.d.M. Saavedra Rios, V. Simone, L. Simonin, S. Martinet, C. Dupont, Biochars
756 from various biomass types as precursors for hard carbon anodes in sodium-ion
757 batteries, *Biomass and Bioenergy*, 117 (2018) 32-37. 10.1016/j.biombioe.2018.07.001.

758 [15] C. Nita, B. Zhang, J. Dentzer, C. Matei Ghimbeu, Hard carbon derived from
759 coconut shells, walnut shells, and corn silk biomass waste exhibiting high capacity for
760 Na-ion batteries, *Journal of Energy Chemistry*, 58 (2021) 207-218.
761 10.1016/j.jechem.2020.08.065.

762 [16] X. Dou, I. Hasa, M. Hekmatfar, T. Diemant, R.J. Behm, D. Buchholz, S. Passerini,
763 Pectin, Hemicellulose, or Lignin? Impact of the Biowaste Source on the Performance of
764 Hard Carbons for Sodium-Ion Batteries, *ChemSusChem*, 10 (2017) 2668-2676.
765 10.1002/cssc.201700628.

766 [17] V. Gupta, V.K. Mishra, P. Singhal, A. Kumar, An Overview of Supervised Machine
767 Learning Algorithm, 2022. 10.1109/SMART55829.2022.10047618.

768 [18] Y. Roh, G. Heo, S.E. Whang, A Survey on Data Collection for Machine Learning: A
769 Big Data - AI Integration Perspective, *IEEE Transactions on Knowledge and Data*
770 *Engineering*, 33 (2021) 1328-1347. 10.1109/tkde.2019.2946162.

771 [19] G.S. Handelman, H.K. Kok, R.V. Chandra, A.H. Razavi, M.J. Lee, H. Asadi,
772 eDoctor: machine learning and the future of medicine, *J Intern Med*, 284 (2018) 603-
773 619. 10.1111/joim.12822.

774 [20] L. Zhang, J. Wen, Y. Li, J. Chen, Y. Ye, Y. Fu, W. Livingood, A review of machine
775 learning in building load prediction, *Applied Energy*, 285 (2021).
776 10.1016/j.apenergy.2021.116452.

777 [21] Y. Liu, B. Guo, X. Zou, Y. Li, S. Shi, Machine learning assisted materials design
778 and discovery for rechargeable batteries, *Energy Storage Materials*, 31 (2020) 434-450.
779 10.1016/j.ensm.2020.06.033.

780 [22] M. Kim, W.-H. Yeo, K. Min, Co-free and low strain cathode materials for sodium-
781 ion batteries: Machine learning-based materials discovery, *Energy Storage Materials*, 69
782 (2024). 10.1016/j.ensm.2024.103405.

783 [23] B. Oral, B. Tekin, D. Eroglu, R. Yildirim, Performance analysis of Na-ion batteries
784 by machine learning, *Journal of Power Sources*, 549 (2022).
785 10.1016/j.jpowsour.2022.232126.

786 [24] V. Soundharrajan, M.H. Alfuruqi, G. Alfaza, J. Lee, S. Lee, S. Park, S. Nithiananth,
787 D.T. Pham, J.-Y. Hwang, J. Kim, Exploring low-cost high energy NASICON cathodes
788 for sodium-ion batteries via a combined machine-learning, ab initio, and experimental
789 approach, *Journal of Materials Chemistry A*, 11 (2023) 15518-15531.
790 10.1039/d3ta02291a.

791 [25] W. Sun, H. Xu, B. Zhou, Y. Guo, Y. Tang, W. Yao, Z. Yang, State-of-charge
792 estimation of sodium-ion batteries: A fusion deep learning approach, *Journal of Energy*
793 *Storage*, 91 (2024). 10.1016/j.est.2024.111527.

794 [26] X. Liu, T. Wang, T. Ji, H. Wang, H. Liu, J. Li, D. Chao, Using machine learning to
795 screen non-graphite carbon materials based on Na-ion storage properties, *Journal of*
796 *Materials Chemistry A*, 10 (2022) 8031-8046. 10.1039/d1ta10588d.

797 [27] T. Ji, X. Liu, D. Sheng, Y. Li, H. Ruan, H. Guo, Z.X. Shen, L. Lai, Machine
798 learning-assisted thermomechanical coupling fabrication of hard carbon for sodium-ion
799 batteries, *Energy Storage Materials*, 71 (2024). 10.1016/j.ensm.2024.103563.

800 [28] S.V. Vassilev, D. Baxter, L.K. Andersen, C.G. Vassileva, T.J. Morgan, An overview
801 of the organic and inorganic phase composition of biomass, *Fuel*, 94 (2012) 1-33.
802 10.1016/j.fuel.2011.09.030.

803 [29] A. Pathy, S. Meher, B. P, Predicting algal biochar yield using eXtreme Gradient
804 Boosting (XGB) algorithm of machine learning methods, *Algal Research*, 50 (2020)
805 102006. 10.1016/j.algal.2020.102006.

806 [30] G. Zhang, C. Chen, C. Xu, J. Li, H. Ye, A. Wang, X. Cao, K. Sun, J. Jiang,
807 Unraveling the Microcrystalline Carbon Evolution Mechanism of Biomass-Derived
808 Hard Carbon for Sodium-Ion Batteries, *Energy & Fuels*, 38 (2024) 8326-8336.
809 10.1021/acs.energyfuels.4c00823.

810 [31] A. Adamson, R. Vali, M. Paalo, J. Aruvali, M. Koppel, R. Palm, E. Hark, J. Nerut,
811 T. Romann, E. Lust, A. Janes, Peat-derived hard carbon electrodes with superior
812 capacity for sodium-ion batteries, *RSC Adv*, 10 (2020) 20145-20154.
813 10.1039/d0ra03212c.

814 [32] M. Vaishnavi, P.M. Vasanth, S. Rajkumar, K.P. Gopinath, Y. Devarajan, A critical
815 review of the correlative effect of process parameters on pyrolysis of plastic wastes,
816 *Journal of Analytical and Applied Pyrolysis*, 170 (2023) 105907.
817 10.1016/j.jaap.2023.105907.

818 [33] F. Güleç, L.M.G. Riesco, O. Williams, E.T. Kostas, A. Samson, E. Lester,
819 Hydrothermal conversion of different lignocellulosic biomass feedstocks – Effect of the
820 process conditions on hydrochar structures, *Fuel*, 302 (2021) 121166.
821 10.1016/j.fuel.2021.121166.

822 [34] C. Matei Ghimbeu, A. Beda, B. Réty, H. El Marouazi, A. Vizintin, B. Tratnik, L.
823 Simonin, J. Michel, J. Abou-Rjeily, R. Dominko, Review: Insights on Hard Carbon
824 Materials for Sodium-Ion Batteries (SIBs): Synthesis – Properties – Performance
825 Relationships, *Advanced Energy Materials*, 14 (2024) 2303833.
826 10.1002/aenm.202303833.

827 [35] C. Wu, Y. Yang, Y. Zhang, H. Xu, X. He, X. Wu, S. Chou, Hard carbon for sodium-
828 ion batteries: progress, strategies and future perspective, *Chem Sci*, 15 (2024) 6244-
829 6268. 10.1039/d4sc00734d.

830 [36] D. Qin, Z. Liu, Y. Zhao, G. Xu, F. Zhang, X. Zhang, A sustainable route from corn
831 stalks to N, P-dual doping carbon sheets toward high performance sodium-ion batteries
832 anode, *Carbon*, 130 (2018) 664-671. 10.1016/j.carbon.2018.01.007.

833 [37] C.H. Tsai, W.T. Tsai, L.A. Kuo, Effect of Post-Washing on Textural Characteristics
834 of Carbon Materials Derived from Pineapple Peel Biomass, *Materials (Basel)*, 16
835 (2023). 10.3390/ma16247529.

836 [38] A. Beda, J.-M. Le Meins, P.-L. Taberna, P. Simon, C. Matei Ghimbeu, Impact of
837 biomass inorganic impurities on hard carbon properties and performance in Na-ion
838 batteries, *Sustainable Materials and Technologies*, 26 (2020).
839 10.1016/j.susmat.2020.e00227.

840 [39] P. Yu, W. Tang, F.-F. Wu, C. Zhang, H.-Y. Luo, H. Liu, Z.-G. Wang, Recent progress
841 in plant-derived hard carbon anode materials for sodium-ion batteries: a review, *Rare*
842 *Metals*, 39 (2020) 1019-1033. 10.1007/s12598-020-01443-z.

843 [40] Q. Jin, K. Wang, H. Li, W. Li, P. Feng, Z. Zhang, W. Wang, M. Zhou, K. Jiang,
844 Tuning microstructures of hard carbon for high capacity and rate sodium storage,
845 *Chemical Engineering Journal*, 417 (2021). 10.1016/j.cej.2020.128104.

846 [41] H.-Y. Yu, H.-J. Liang, Z.-Y. Gu, Y.-F. Meng, M. Yang, M.-X. Yu, C.-D. Zhao, X.-L.
847 Wu, Waste-to-wealth: low-cost hard carbon anode derived from unburned charcoal with
848 high capacity and long cycle life for sodium-ion/lithium-ion batteries, *Electrochimica*
849 *Acta*, 361 (2020). 10.1016/j.electacta.2020.137041.

850

851