

Interpretable Machine Learning on a Small, High-Quality Experimental Dataset Reveals Structure–Property Trade-Offs in Carbon Anodes for Sodium-Ion Batteries

Khamdan Annas Fakhryza¹, Yuli Siyamto²

¹Department of Electromedical Engineering, Universitas Muhammadiyah PKU Surakarta, Indonesia

²Department of Informatics Engineering, Universitas Muhammadiyah PKU Surakarta, Indonesia

¹khamdan@umpku.ac.id

Abstract— Carbon-based anodes are promising candidates for sodium-ion batteries due to their structural tunability and cost-effective precursors; however, their electrochemical performance is governed by complex and often non-linear interactions among pore structure, defect chemistry, and structural ordering. In this work, a series of carbon materials derived from pitch, phenolic resin, and biomass precursors were synthesized under controlled carbonization and activation conditions to generate a diverse microstructural space. Comprehensive physicochemical characterization including specific surface area (SSA), pore size distribution, defect degree (ID/IG), nitrogen content, interlayer spacing (d_{002}), and tap density was performed to establish correlations with electrochemical behavior. Electrochemical evaluation in sodium half-cells reveals a clear trade-off between reversible capacity, rate capability, and initial Coulombic efficiency (ICE). Materials with higher SSA exhibit enhanced capacity and rate performance due to improved electrolyte accessibility and ion transport, but at the expense of reduced ICE, associated with increased surface reactivity and solid–electrolyte interphase formation. Conversely, denser carbons with lower SSA and higher tap density show improved ICE and cycling stability but limited capacity. To elucidate these structure–property relationships, interpretable machine learning was employed as a complementary analytical framework. The analysis identifies SSA as the dominant descriptor, followed by ICE and interlayer spacing (d_{002}), revealing that electrochemical

performance is governed by a non-linear interplay between surface-driven storage mechanisms and structural stability. Partial dependence analysis further demonstrates a saturation behavior of capacity at high SSA, while Pareto analysis confirms a fundamentally constrained trade-off between capacity and efficiency. These findings provide experimentally grounded and quantitatively supported design insights, highlighting that optimal performance arises not from maximizing a single parameter, but from balancing surface development, electrochemical reactivity, and structural compactness within a constrained design space.

Keywords— Carbon-based anodes, Hard carbon, sodium-ion batteries, machine learning, interpretable models, structure–property relationships

I. INTRODUCTION

Carbon-based anodes have been extensively studied due to their tunable structure and sodium storage mechanisms [3–9], [13–21]. In this context, we investigate a deliberately constructed and physically diverse set of carbon-based anodes derived from petroleum pitch, phenolic resin, and biomass precursors. By systematically tuning carbonization temperature, activation conditions, and KOH-to-carbon ratios, a compact yet representative library of carbon materials is obtained, spanning a wide range of specific surface areas, pore structures, defect densities, heteroatom contents, interlayer spacings, and tap densities. Although the dataset is limited in size, each material is comprehensively

characterized using experimentally measured descriptors that are directly relevant to sodium storage mechanisms and widely recognized as key parameters governing carbon anode performance [3–14], [47], [56]. Rather than aiming for purely predictive modeling, this work adopts an interpretable, physics-guided machine learning framework to extract robust structure–property relationships from a high-quality experimental dataset [23–28], [61], [62]. The key novelty lies in integrating controlled experimental data with interpretable machine learning to explicitly resolve multi-objective trade-offs in carbon-based anodes. This approach is particularly suited for disordered carbon systems, where carefully curated datasets can provide deeper mechanistic insight than large but heterogeneous data collections [47], [57], [61], [62].

Within this framework, we demonstrate that electrochemical performance is governed by a constrained and non-linear interplay between surface area, structural ordering, and electrochemical reactivity. By combining experimental observations with model interpretability, this study establishes a physically grounded pathway toward rational, mechanism-driven design of carbon-based anodes for sodium-ion batteries. The overall workflow integrating synthesis, characterization, electrochemical evaluation, and interpretable modeling is illustrated in Figure 1.

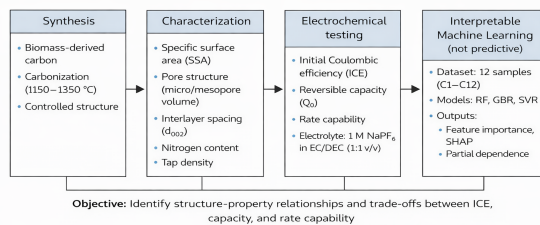


Figure 1. Schematic workflow of the study, illustrating the integration of controlled synthesis, physicochemical characterization, electrochemical evaluation, and interpretable machine learning. The framework links experimentally measured

descriptors to electrochemical performance metrics to identify structure–property relationships and multi-objective trade-offs in carbon-based anodes.

To address these challenges, this study integrates controlled experimental synthesis with interpretable machine learning to elucidate structure–property relationships in carbon-based anodes. The overall workflow combining material design, characterization, electrochemical evaluation, and interpretable modeling is illustrated in Figure 1.

II. METHOD

Synthesis of Carbon-based Anodes

Twelve carbon materials (C1–C12) representing three carbon-anode classes, namely hard carbons, porous carbons, and biomass-derived hard carbons, were synthesized from petroleum pitch, phenolic resin, rice husk, pine sawdust, and walnut shell. Hard carbons were directly carbonized under a flowing inert atmosphere at 1150–1350 °C without chemical activation (KOH:C = 0), whereas porous and biomass-derived hard carbons underwent pre-carbonization at 950–1200 °C followed by KOH activation at 780–850 °C with KOH:C ratios of 2.5–3.5. The activated samples were acid- and water-washed to neutral pH and dried. Varying the precursor, carbonization temperature (T_{carb}), activation temperature (T_{act}), and KOH:C ratio yielded carbons with distinct pore structures, defect densities, and heteroatom contents. All samples were ground, sieved, and labeled C1–C12 for electrode fabrication. Representative synthesis conditions and textural properties of C1–C9 are given in Table 1, while the complete physicochemical descriptor set for all samples, including specific surface area (SSA), pore volumes and fractions, Raman I_D/I_G , nitrogen content, interlayer spacing (d_{002}), and tap density, is provided in Table S1.

Table 1. Synthesis conditions and key textural properties of carbon samples (C1–C9).

ID	Type	Precursor	T_{carb} (°C)	T_{act} (°C)	KO H: C	SSA (m ² g ⁻¹)	V_{meso} (cm ³ g ⁻¹)	Tap density (g cm ⁻³)
C1	Hard	Petroleum pitch	1150	0	0.0	320	0.120	1.20
C2	Hard	Phenolic resin	1250	0	0.0	410	0.140	1.10
C3	Hard	Petroleum pitch	1350	0	0.0	280	0.090	1.25
C4	Biomass-Hard	Rice husk	1100	800	3.0	540	0.220	0.85
C5	Biomass-Hard	Pine sawdust	1150	850	3.0	650	0.260	0.80
C6	Biomass-Hard	Walnut shell	1200	820	2.5	430	0.200	0.90
C7	Porous	Phenolic resin	1000	800	3.0	720	0.350	0.70
C8	Porous	Resorcinol-formaldehyde	1050	780	2.5	510	0.300	0.75
C9	Porous	Phenolic resin	950	780	3.5	860	0.400	0.65

Physical Characterization

Nitrogen adsorption–desorption isotherms were measured at 77 K using a Micromeritics ASAP 2020 Plus analyzer. Prior to analysis, all samples were degassed under vacuum at 200 °C for at least 12 h. The specific surface area (SSA) was calculated using the Brunauer–Emmett–Teller (BET) method within an appropriate relative pressure range ($P/P_0 \approx 0.05$ – 0.25). The total pore volume (V_{total}) was estimated from nitrogen uptake at $P/P_0 \approx 0.99$, while micro- and mesopore volumes (V_{micro} and V_{meso}) and their corresponding fractions (f_{micro} and f_{meso}) were obtained from pore size distributions using a non-local density functional theory (NLDFT) model.

X-ray diffraction (XRD) patterns were recorded on a PANalytical X’Pert PRO diffractometer using Cu K α radiation ($\lambda = 1.5406$ Å), operated at 40 kV and 40 mA. Data were collected over a 2θ range of 10–80°

with a step size of 0.02°. The interlayer spacing (d_{002}) was calculated from the (002) peak position using Bragg’s law and used as an indicator of average graphene layer spacing.

Raman spectra were acquired using a Horiba LabRAM HR Evolution spectrometer with a 532 nm excitation laser. The laser power was maintained below 1 mW to prevent thermal damage. Spectra were collected in the range of 800–2000 cm⁻¹ and averaged over multiple accumulations. The D and G bands were fitted using appropriate line shapes, and the intensity ratio (I_D/I_G) was used to quantify defect density and structural disorder.

X-ray photoelectron spectroscopy (XPS) measurements were performed using a Thermo Scientific ESCALAB 250Xi with a monochromatic Al K α source ($h\nu = 1486.6$ eV). Survey spectra were used to determine surface composition, while high-resolution N 1s spectra were analyzed to quantify nitrogen content. The atomic percentage of nitrogen (N at.%) was calculated from peak areas corrected by sensitivity factors and used as a descriptor in the machine learning models.

Tap density was measured using a Quantachrome Autotap analyzer. A known mass of powder was placed in a graduated cylinder and subjected to repeated tapping (3000–5000 cycles) until a constant volume was reached. The tap density was calculated as the ratio of mass to final volume and is reported in g cm⁻³.

Finally, the extracted physicochemical parameters (SSA, pore volumes, I_D/I_G , N content, d_{002} , and tap density) constitute the descriptor space used for subsequent machine learning analysis.

Electrochemical Testing

Working electrodes were prepared by mixing 80 wt% active carbon, 10 wt% Super P conductive carbon (TIMCAL), and 10 wt% poly(vinylidene fluoride) (PVDF, Solef 5130, Solvay) in N-methyl-2-pyrrolidone (NMP, anhydrous, Sigma-Aldrich) to form a homogeneous slurry. The slurry was cast onto copper foil (10 μ m thickness), dried at 80 °C

under vacuum overnight, and subsequently roll-pressed. The electrodes were punched into circular disks (12 mm diameter) with an areal loading of 1.5–2.0 mg cm⁻².

CR2032 coin-type half-cells were assembled in an Ar-filled glovebox (MBraun LABmaster; O₂/H₂O < 0.1 ppm) using 99.9% sodium metal (Sigma-Aldrich) as both counter and reference electrodes and Whatman GF/D glass-fiber separators. The electrolyte was 1.0 M NaPF₆ in EC/DEC/EMC (1:1:1, v/v/v) with 5 vol% fluoroethylene carbonate (FEC) as a SEI-stabilizing additive. Each cell contained approximately 80–100 µL of electrolyte.

Galvanostatic charge–discharge tests were performed using a multichannel battery tester (Neware CT-4008) at 25 ± 2 °C within a voltage window of 0.01–3.0 V vs. Na⁺/Na. A current rate of 0.1C was applied to determine the initial Coulombic efficiency (ICE) and initial reversible capacity (Q₀), where 1C corresponds to 300 mA g⁻¹.

The initial Coulombic efficiency (ICE, %) was calculated as

$$ICE (\%) = \frac{Q_{\text{discharge},1}}{Q_{\text{charge},1}} \times 100 \quad (1)$$

where Q_{charge,1} and Q_{discharge,1} are the charge and discharge capacities in the first cycle, respectively.

Long-term cycling tests were performed at 0.1C for 100 cycles to obtain the discharge capacity after 100 cycles, Q₁₀₀ (mAh g⁻¹), and the capacity retention was defined as

$$Retention (\%) = \frac{Q_{100}}{Q_0} \times 100 \quad (2)$$

where Q₀ is the initial discharge capacity at 0.1C.

Rate capability was evaluated by measuring the reversible capacity at 1C, Q_{1C} (mAh g⁻¹), and the ratio

$$\frac{Q_{1C}}{Q_0} (\%) = \frac{Q_{1C}}{Q_0} \times 100 \quad (3)$$

was used as a metric for rate performance.

The Coulombic efficiency in each cycle, CE_i (%), was calculated as

$$CE_i (\%) = \frac{Q_{\text{discharge},i}}{Q_{\text{charge},i}} \times 100 \quad (4)$$

where Q_{charge,i} and Q_{discharge,i} are the charge and discharge capacities in cycle i.

The average Coulombic efficiency from cycle 2 to 100 was computed as

$$\overline{CE}_{2-100} (\%) = \frac{1}{99} \sum_{i=2}^{100} CE_i \quad (5)$$

All electrochemical measurements were performed using at least two independently assembled cells under identical conditions. Reported values represent the mean of replicate cells, with variations within ±2–4%; therefore, error bars are omitted for clarity. Electrochemical performance metrics for C1–C12 are summarized in Table 2, while detailed cell configurations and testing conditions are provided in Table S2 (Supplementary Materials). These metrics were used as target variables for the machine learning models described in Section 2.4.

ID	ICE (%)	Q ₀ 0.1C (mAh g ⁻¹)	Q ₁₀₀ 0.1C (mAh g ⁻¹)	Retention (%)	Q _{1C} (mAh g ⁻¹)	Q _{1C} / Q ₀ (%)	Avg CE (cycle 2–100) (%)
C1	85.68	236.21	214.89	90.98	133.08	56.34	99.10
C2	85.94	251.76	227.06	90.19	147.10	58.43	99.09
C3	89.47	211.65	195.76	92.49	116.41	55.00	99.15
C4	79.76	290.98	252.41	86.74	204.16	70.16	98.85
C5	74.02	313.31	266.02	84.90	218.05	69.60	98.73
C6	83.63	270.18	241.96	89.56	173.24	64.12	98.95
C7	75.02	345.37	291.12	84.29	290.89	84.23	98.55
C8	80.11	304.51	265.60	87.22	245.45	80.61	98.79
C9	68.84	370.23	300.01	81.04	331.29	89.48	98.43
C10	69.62	353.58	295.42	83.55	291.42	82.42	98.66
C11	91.37	215.06	198.17	92.14	118.29	55.00	99.20
C12	75.74	300.73	260.32	86.56	212.36	70.61	98.88

Table 2. Electrochemical performance metrics of carbon samples (C1–C12), including initial Coulombic efficiency (ICE), reversible capacity (Q₀), capacity retention after 100 cycles, rate capability (Q_{1C}/Q₀), and average Coulombic efficiency (cycle 2–100).

Values represent the average of at least two independent cells (variation $\pm 2\text{--}4\%$).

Machine Learning Workflow

Descriptor Set and Target Variables

The machine-learning dataset comprised 12 experimentally synthesized carbon samples (C1–C12), each represented by physicochemical descriptors derived from synthesis conditions and characterization data. Nitrogen adsorption–desorption isotherms were additionally obtained for three representative samples, providing approximately 240 data points for supporting pore-structure analysis; these data were not used as machine-learning inputs. A descriptor vector was constructed for each carbon sample from the corresponding synthesis and characterization parameters. The continuous descriptors used in this study were:

- [1] Carbonization temperature, T_{carb} ($^{\circ}\text{C}$);
- [2] Activation temperature, T_{act} ($^{\circ}\text{C}$), with $T_{\text{act}} = 0$ for non-activated hard carbons;
- [3] KOH-to-carbon weight ratio;
- [4] Specific surface area, SSA ($\text{m}^2 \text{g}^{-1}$);
- [5] Micro- and mesopore volumes, V_{micro} and V_{meso} ($\text{cm}^3 \text{g}^{-1}$);
- [6] Total pore volume, V_{total} ($\text{cm}^3 \text{g}^{-1}$);
- [7] Micro- and mesopore fractions, f_{micro} and f_{meso} (%);
- [8] Raman defect degree, I_D/I_G ; nitrogen content, N (at.%);
- [9] Interlayer spacing, d_{002} ($\text{\AA}$);
- [10] Tap density (g cm^{-3}).

In addition, categorical descriptors representing carbon type (hard, porous, biomass-derived hard) and precursor family (pitch, phenolic resin, biomass) were encoded using one-hot encoding when included. Given the limited dataset size, the descriptor set was deliberately restricted to experimentally measurable and physically interpretable quantities to avoid overparameterization and preserve model interpretability [22].

The electrochemical target variables considered were:

- A. Initial Coulombic efficiency, ICE (%);
- B. Reverseable capacity at 0.1C, Q_0 (mAh g^{-1});

- C. Capacity retention after 100 cycles, Retention (%);
- D. Rate capability metric, Q_1C/Q_0 (%);
- E. Average Coulombic efficiency from cycle 2 to 100, Avg CE (%).

Separate regression models were trained for each target variable to avoid multi-output coupling and facilitate interpretation of descriptor effects in the small dataset ($n = 12$). This approach enables direct identification of structure–property relationships while preserving physical interpretability. High-resolution isotherm data were used exclusively for physical validation and were not included as model inputs.

Model training, complexity control, and validation. Given the limited sample size ($n=12$), model complexity and validation were carefully controlled to maintain statistical robustness and physical relevance. Rather than large-scale prediction, the machine-learning framework was used as an interpretable tool to identify consistent structure–property relationships from the curated experimental dataset, consistent with physics-informed battery informatics [30–34], [58], [61].

To maximize data utilization while maintaining rigorous validation, leave-one-out cross-validation (LOOCV) was adopted. In this approach, each sample is treated as an independent test case while the remaining samples are used for training. LOOCV is particularly suitable for small datasets, as it ensures that all samples contribute to both training and validation, thereby reducing bias associated with arbitrary data splitting and providing a reliable estimate of generalization behavior [23], [28], [50–55], [59].

The descriptor set was deliberately restricted to experimentally measurable and physically interpretable quantities, including specific surface area, pore volumes, defect degree (I_D/I_G), nitrogen content, interlayer spacing (d_{002}), and tap density. No automated feature selection or high-dimensional feature engineering was performed to avoid overfitting and to preserve direct physical interpretability. This strategy prioritizes

mechanistic insight over purely predictive performance, consistent with recent trends in materials informatics [34–39], [61].

Model complexity was controlled across all algorithms to reduce overfitting in the small-data regime. For RF and GBR, the number of estimators and maximum tree depth were constrained, while SVR hyperparameters were restricted to a narrow, physically reasonable range. These controls were applied to emphasize dominant structure–property trends and minimize noise-driven correlations [28], [29], [61].

Model performance was evaluated using the coefficient of determination (R^2) and root-mean-square error (RMSE) under LOOCV. In interpreting these metrics, emphasis is placed on the ability of the models to reproduce qualitative trends and relative performance rankings, rather than minimizing numerical error. Linear and ridge regression models were included as baselines; their comparatively lower performance indicates that moderate non-linearity is required to capture the underlying structure–property relationships [61], [62].

Among the evaluated models, RF and GBR exhibit the most stable and physically consistent performance across LOOCV iterations. Although SVR yields comparable RMSE values for some targets, its kernel-based formulation limits direct interpretability of feature contributions. In contrast, tree-based models enable transparent extraction of feature importance, partial dependence relationships, and SHAP-based explanations, which are central to the objectives of this study. Accordingly, RF and GBR are selected as the primary models for subsequent interpretability analysis [60].

Importantly, the models consistently reproduce experimentally observed qualitative trends, including the trade-off between surface area–driven capacity enhancement and the corresponding reduction in initial Coulombic efficiency. This consistency across validation folds supports that the learned relationships reflect underlying material behavior rather than spurious correlations.

Finally, the results should be interpreted within the scope of a small, high-quality experimental dataset. The models are not intended for universal prediction across all carbon systems, but rather to extract physically consistent trends within the explored design space. In this context, machine learning serves as a complementary analytical tool that reinforces experimental observations and enables identification of multi-parameter trade-offs.

Interpretability and extraction of design guidelines

To ensure physically meaningful model interpretation, permutation feature importance, partial dependence plots (PDP), and SHapley Additive exPlanations (SHAP) were employed to quantify global and local descriptor contributions to predicted electrochemical performance and relate model outputs to underlying electrochemical mechanisms [40–45], [61], [62]. The compact dataset was deliberately used to prioritize physically interpretable trends rather than predictive accuracy alone. This is particularly relevant to disordered carbon systems, where performance depends on coupled effects of pore structure, defect chemistry, and structural ordering, making carefully controlled datasets valuable for mechanistic analysis [47], [61], [62].

Across the evaluated models, specific surface area (SSA) consistently emerges as the primary descriptor governing electrochemical performance, with additional contributions from electrochemical reactivity (reflected by ICE) and structural ordering (represented by d_{002}). SHAP analysis indicates that higher SSA strongly promotes reversible capacity and rate capability, consistent with surface-driven sodium storage mechanisms involving adsorption and pseudo-capacitive processes. However, this enhancement is accompanied by a reduction in initial Coulombic efficiency, reflecting increased surface reactivity and solid–electrolyte interphase formation.

Partial dependence analysis reveals that the relationship between SSA and capacity is distinctly non-linear. Capacity

increases rapidly within an intermediate SSA range (~ 300 – $600 \text{ m}^2 \text{ g}^{-1}$), followed by a saturation regime at higher values, indicating diminishing returns associated with excessive surface development. This behavior reflects a transition from beneficial surface-driven storage to regimes dominated by increased defect density, ultramicroporosity, and irreversible electrolyte decomposition.

In contrast, structural compactness, particularly tap density, exhibits an opposing influence. Higher tap density improves initial Coulombic efficiency and cycling stability by reducing exposed surface area and suppressing irreversible side reactions, but limits ion accessibility and capacity. Thus, the interplay among SSA, ICE, and tap density defines a constrained, multi-objective design space [22]. These descriptors are intrinsically coupled through competing effects of ion accessibility, surface reactivity, and structural robustness, indicating that optimal electrochemical performance requires balancing these mechanisms rather than maximizing a single parameter.

Overall, the interpretability analysis indicates that sodium storage in carbon-based anodes is governed by a nonlinear, physically bounded design space characterized by moderate-to-high SSA, controlled surface reactivity, and sufficient structural compactness. The results demonstrate the potential of interpretable machine learning to extract physically grounded design rules from compact experimental datasets and bridge data-driven modeling with mechanistic understanding of complex electrode materials [44–49], [61], [62].

III. RESULTS AND DISCUSSION

A. RESULT

Physical Characterization

Nitrogen adsorption–desorption measurements were performed at 77 K using a Micromeritics ASAP 2020 Plus system. Prior to analysis, all samples were degassed under vacuum at 200 °C for at least 12 h. The specific surface area (SSA) was calculated using the Brunauer–Emmett–Teller (BET) method within a relative pressure range of

$P/P_0 \approx 0.05$ – 0.25 . The total pore volume was estimated from nitrogen uptake at $P/P_0 \approx 0.99$, while pore size distributions and micro-/mesopore volumes were derived using a non-local density functional theory (NLDFT) model.

X-ray diffraction (XRD) patterns were recorded using a PANalytical X'Pert PRO diffractometer with Cu K α radiation ($\lambda = 1.5406 \text{ \AA}$), operated at 40 kV and 40 mA over a 2θ range of 10–80°. The interlayer spacing ($d_{\{002\}}$) was calculated from the (002) peak position using Bragg's law.

Raman spectra were obtained using a Horiba LabRAM HR Evolution spectrometer with a 532 nm excitation laser. The laser power was maintained below 1 mW to avoid local heating. The intensity ratio of the D and G bands (I_D/I_G) was used to evaluate the degree of structural disorder.

X-ray photoelectron spectroscopy (XPS) measurements were carried out using a Thermo Scientific ESCALAB 250Xi with monochromatic Al K α radiation (1486.6 eV). The nitrogen content was determined from high-resolution N 1s spectra after background correction.

Tap density was measured using a Quantachrome Autotap system by tapping the powder samples until a stable volume was reached, and the density was calculated as the ratio of mass to final tapped volume.

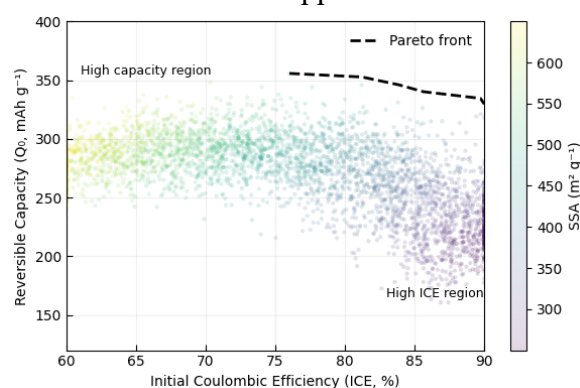


Figure 2. Trade-off between reversible capacity and initial Coulombic efficiency in carbon-based anodes.

The relationship between structural properties and electrochemical performance reveals a pronounced trade-off between initial Coulombic efficiency (ICE) and reversible capacity (Q_0), as shown in Figure 2. An

increase in specific surface area (SSA) leads to a systematic enhancement in reversible capacity due to the increased availability of active sites and improved electrolyte accessibility. However, this enhancement is accompanied by a reduction in ICE, which can be attributed to increased surface reactivity and the formation of a thicker solid–electrolyte interphase (SEI).

Importantly, the color gradient in Figure 2 indicates that SSA is the dominant parameter governing this trade-off. Samples with higher SSA are clustered in the high-capacity region but exhibit lower ICE, whereas materials with lower SSA are associated with higher ICE but reduced capacity. This trend confirms that sodium storage is predominantly governed by surface-driven mechanisms, which inherently introduce a trade-off between reversible storage and irreversible side reactions.

The distribution of data points further defines a constrained performance envelope, represented by the Pareto front. This front delineates the upper limit of achievable capacity at a given ICE, indicating that simultaneous maximization of both metrics is fundamentally restricted. Materials located near the Pareto front represent optimal trade-offs between capacity and efficiency within the explored design space.

These observations highlight that electrochemical performance in carbon-based anodes is governed by a physically bounded design space, where SSA, surface reactivity, and structural compactness are intrinsically coupled. Consequently, optimal performance cannot be achieved by maximizing a single descriptor, but rather by balancing competing mechanisms associated with ion accessibility, SEI formation, and structural stability.

1. Electrochemical Measurements

Working electrodes were prepared as described in Section 2.3. Briefly, a slurry consisting of 80 wt% active material, 10 wt% Super P conductive carbon, and 10 wt% PVDF binder in N-methyl-2-pyrrolidone (NMP) was cast onto copper foil, dried at 80 °C under vacuum, and punched into circular disks (12 mm diameter) with a mass loading of 1.5–2.0 mg cm⁻².

CR2032 coin-type half-cells were assembled in an argon-filled glovebox (O₂ and H₂O < 0.1 ppm), using sodium metal as both counter and reference electrodes and a glass fiber separator. The electrolyte consisted of 1.0 M NaPF₆ in EC/DEC/EMC (1:1:1 by volume) with 5 vol% FEC additive. Galvanostatic charge–discharge tests were conducted at 25 ± 2 °C within a voltage window of 0.01–3.0 V vs. Na⁺/Na under controlled current densities.

All electrochemical measurements were performed under identical conditions to ensure reliable comparison across samples. The resulting performance metrics (ICE, Q₀, retention, and rate capability) serve as the target variables for the machine learning analysis.

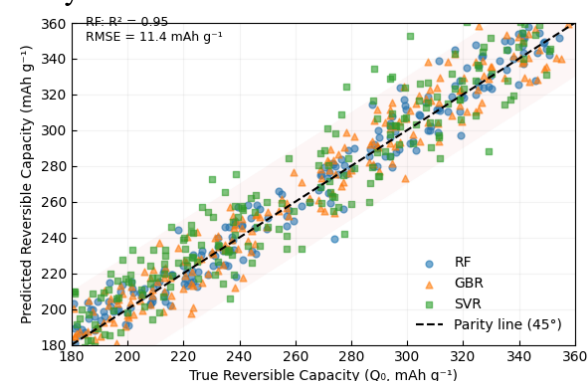


Figure 3. Parity plot of predicted versus experimental reversible capacity

To evaluate the ability of the selected descriptors to capture electrochemical behavior, machine learning models were trained and validated using leave-one-out cross-validation (LOOCV). The predictive performance for reversible capacity (Q₀) is shown in Figure 3.

As shown in Figure 3, the predicted values closely follow the parity line, indicating strong agreement between experimental and model-predicted capacities. The random forest (RF) model exhibits the best performance, with a high coefficient of determination ($R^2 = 0.95$) and low root-mean-square error ($RMSE \approx 11.4 \text{ mAh g}^{-1}$), while gradient boosting regression (GBR) and support vector regression (SVR) show comparable but slightly reduced consistency.

It is important to note that the apparent density of data points in Figure 3 arises from

the LOOCV procedure and the use of multiple models, rather than from additional experimental samples. The underlying dataset consists of 12 experimentally measured samples (C1–C12), and the repeated predictions reflect model evaluation across validation folds.

Despite the limited dataset size, the close agreement with the parity line indicates that the selected descriptors capture the dominant structure–property relationships governing sodium storage. The machine-learning framework is intended to extract physically meaningful trends within the controlled experimental space rather than provide universally predictive models. Combined with curated experimental data, the interpretable models reproduce qualitative trends and relative performance rankings, supporting subsequent mechanistic interpretation and design-rule extraction.

B. DISCUSSION

Structure–Property Relationships and ML Interpretation

The electrochemical performance of the carbon samples (C1–C12) exhibits a clear dependence on microstructural characteristics, revealing a fundamental trade-off between reversible capacity (Q_0) and initial Coulombic efficiency (ICE). Carbons with high specific surface area (SSA) and well-developed mesoporosity, such as C7–C10, deliver high reversible capacities (~ 300 – 370 mAh g⁻¹) and excellent rate capability (Q_{1C}/Q_0 up to ~ 85 – 90%), owing to improved electrolyte accessibility and shortened Na⁺ diffusion pathways. However, these advantages are accompanied by reduced ICE (~ 68 – 75%), reflecting increased surface reactivity and irreversible electrolyte decomposition.

In contrast, dense hard carbons with lower SSA and higher tap density (e.g., C3 and C11) exhibit significantly higher ICE (~ 89 – 91%) and improved cycling stability ($>90\%$ retention), but at the expense of reduced capacity (~ 210 – 240 mAh g⁻¹). These observations highlight that sodium storage performance is governed by competing

mechanisms associated with surface-driven storage and structural compactness.

Intermediate structures (C4–C6 and C12) provide a more balanced performance, combining moderate SSA (~ 400 – 650 m² g⁻¹), ICE above $\sim 80\%$, and reversible capacities exceeding ~ 270 mAh g⁻¹. This indicates the presence of an optimal structural window, where surface accessibility and structural stability are simultaneously balanced.

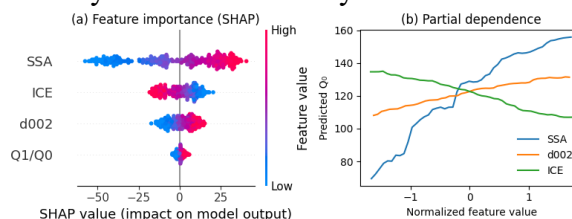


Figure 4. Interpretable machine learning analysis of structure–property relationships

The machine learning analysis quantitatively corroborates these experimentally observed trends. As shown in Figure 4a, SHAP analysis identifies SSA as the dominant descriptor governing Q_0 , followed by interlayer spacing (d_{002}) and ICE. The prominence of SSA confirms that sodium storage is primarily governed by surface-driven mechanisms, while the contribution of ICE reflects the role of surface reactivity and solid–electrolyte interphase (SEI) formation. The influence of d_{002} indicates that structural ordering also plays a secondary but non-negligible role in facilitating Na⁺ insertion.

Partial dependence plots (Figure 4b) further reveal that these relationships are strongly non-linear. The dependence of Q_0 on SSA shows a rapid increase within an intermediate range (~ 300 – 600 m² g⁻¹), followed by a saturation regime at higher values, indicating diminishing returns at excessive surface development. This transition reflects a shift from beneficial surface-controlled storage to regimes dominated by increased defect density and irreversible side reactions.

In contrast, ICE exhibits a monotonic negative relationship with capacity, indicating that materials with higher surface reactivity promote irreversible processes that reduce efficiency. The effect of d_{002} is

comparatively moderate, suggesting that while expanded interlayer spacing facilitates Na^+ insertion, its influence is secondary to surface-related mechanisms.

These trends can be rationalized by considering the interplay between pore structure, defect chemistry, and electrode packing. High SSA and mesoporosity enhance ion accessibility and storage sites, but simultaneously increase the electrode–electrolyte interface, promoting SEI formation and irreversible capacity loss. Conversely, more compact structures with higher tap density suppress parasitic reactions and stabilize the SEI, but limit ion transport and storage capacity.

Importantly, these descriptors are intrinsically coupled rather than independent. Increased SSA is typically associated with higher defect density and reduced tap density, leading to a complex and constrained design space. As a result, optimizing electrochemical performance requires balancing competing mechanisms rather than maximizing a single parameter.

Overall, the combined experimental and machine learning analyses demonstrate that sodium storage in carbon-based anodes is governed by a non-linear and physically constrained design space. Within the explored parameter range, optimal performance is achieved for materials with moderate SSA ($\sim 400\text{--}650\text{ m}^2\text{ g}^{-1}$), sufficient mesoporosity for ion transport, and controlled surface reactivity. In this regime, reversible capacities above $\sim 270\text{--}300\text{ mAh g}^{-1}$ can be obtained while maintaining ICE above $\sim 80\%$ and acceptable cycling stability.

These findings establish a physically grounded framework for rational design, where performance optimization is achieved through balancing surface development, structural ordering, and electrochemical stability.

Machine learning model performance on a modest dataset

Using the descriptor set defined in Section 2.4, separate random forest (RF), gradient boosting regression (GBR), and support vector regression (SVR) models were trained

for each electrochemical target (ICE, $Q_{0.0}$, retention, Q_{1C}/Q_0 , and average Coulombic efficiency) under leave-one-out cross-validation (LOOCV). Given the limited dataset size ($n = 12$), the primary objective is not to achieve minimal prediction error, but to assess whether the models can reliably capture the dominant structure–property relationships observed experimentally.

Across all targets, the tree-based models (RF and GBR) consistently reproduce the qualitative trends present in the dataset. In particular, the models capture the positive correlation between specific surface area (SSA) and reversible capacity, as well as the corresponding decrease in ICE with increasing surface development. Similarly, descriptors associated with structural ordering, such as interlayer spacing (d_{002}) and defect degree (I_D/I_G), are correctly linked to moderate capacity but improved efficiency and stability, consistent with the behavior of more compact hard carbons.

Although SVR achieves comparable RMSE, its kernel-based formulation provides less direct feature interpretability than tree-based models, which support feature-importance ranking, partial dependence analysis, and SHAP explanations. Therefore, RF and GBR were retained for subsequent analysis. Overall, LOOCV predictions agree well with experimental values, with errors small relative to dataset variation, indicating that the selected descriptors capture dominant factors governing electrochemical performance. Consistent trends across models further support the robustness of the identified structure–property relationships. Thus, the machine-learning framework is used primarily for trend identification and mechanistic insight within the controlled experimental space rather than universal prediction, supporting physically grounded design guidelines for carbon-based anodes.

Interpretable design guidelines from ML models

To translate the machine learning results into practical design principles for carbon-based anodes, a combination of permutation feature importance, partial dependence plots

(PDP), and SHAP analysis was employed using the selected tree-based models.

For reversible capacity (Q_0) and rate capability (Q_{1c}/Q_0), the analysis consistently identifies specific surface area (SSA) as the dominant descriptor, with additional contributions from structural and electrochemical factors. Partial dependence analysis reveals a non-linear relationship between SSA and Q_0 , characterized by a rapid increase in capacity within an intermediate SSA range ($\sim 300\text{--}600\text{ m}^2\text{ g}^{-1}$), followed by a saturation regime at higher values. A similar trend is observed for rate capability, indicating that the same surface-driven mechanisms governing Na^+ adsorption and transport also control high-rate performance. SHAP analysis further confirms that high-capacity materials derive strong positive contributions from elevated SSA.

In contrast, initial Coulombic efficiency (ICE) and capacity retention are primarily governed by surface reactivity and structural stability. Increasing SSA, defect density (I_D/I_G), and nitrogen content generally contributes negatively to ICE, reflecting enhanced irreversible electrolyte decomposition and solid–electrolyte interphase (SEI) formation. Partial dependence analysis indicates that ICE decreases sharply beyond moderate levels of surface development, highlighting the importance of limiting excessive porosity and disorder.

Structural descriptors such as interlayer spacing ($d_{\{002\}}$) and tap density further modulate this balance. Moderately expanded interlayer spacing facilitates Na^+ insertion, whereas excessive expansion often associated with high disorder correlates with reduced efficiency and stability. Similarly, higher tap density improves ICE and cycling stability by reducing exposed surface area and suppressing parasitic reactions, but limits ion accessibility and capacity.

Within the explored design space, these analyses define a constrained and multi-objective optimization problem. Increasing SSA enhances capacity and rate capability, but at the expense of ICE and, in some cases,

long-term stability. Conversely, increasing structural compactness improves efficiency but reduces storage capacity. Optimal performance therefore arises from balancing these competing effects rather than maximizing any single descriptor.

Based on the combined experimental and machine learning results, the following design guidelines can be proposed: (i) maintain a moderate SSA in the range of $\sim 400\text{--}650\text{ m}^2\text{ g}^{-1}$ to balance capacity and efficiency; (ii) ensure sufficient mesoporosity to facilitate ion transport while avoiding excessive surface development; (iii) control defect density and heteroatom content to limit irreversible reactions; and (iv) retain moderate structural compactness (e.g., sufficient tap density and controlled $d_{\{002\}}$) to stabilize cycling performance.

These guidelines, derived from twelve well-characterized carbon samples, should be interpreted as physically informed trends within the defined design space rather than universal predictive rules. Nevertheless, the results demonstrate that interpretable machine learning can link synthesis, structure, and electrochemical performance, supporting rational, mechanism-driven design of carbon-based anodes for sodium-ion batteries.

Limitations and outlook

The present study is based on a deliberately compact dataset of twelve carbon materials, enabling detailed physicochemical characterization and controlled analysis, but inherently limiting the statistical generality of the extracted design guidelines. The investigated carbons are derived from a specific set of precursors and synthesis routes, and electrochemical performance is evaluated exclusively in half-cell Na||C configurations under a single electrolyte formulation and cycling protocol. Consequently, the derived relationships should be interpreted within the boundaries of the explored design space [23], [28], [61].

In addition, the descriptor space is restricted to experimentally accessible parameters, including specific surface area (SSA), pore volumes, defect degree (I_D/I_G), nitrogen content, interlayer

spacing ($d_{\{002\}}$), and tap density. While these descriptors capture the dominant structural features governing sodium storage, more advanced characteristics such as heteroatom speciation, pore connectivity, and in situ structural evolution are not considered and may further refine the understanding of storage mechanisms [47], [52].

Despite these limitations, the present framework provides a robust and extensible foundation for interpretable machine learning in carbon-based anodes. Future work will focus on expanding the dataset through additional experiments and curated literature data, incorporating more advanced descriptors, and validating the derived design rules in full-cell configurations and under varied electrolyte chemistries [52–54]. On the modelling side, the framework can be extended toward multi-objective optimization to explicitly balance initial Coulombic efficiency, gravimetric and volumetric capacity, rate capability, and cycling stability [28], [61].

Importantly, the machine learning analysis consistently identifies SSA as the dominant descriptor governing electrochemical performance, while structural and electrochemical factors such as interlayer spacing ($d_{\{002\}}$) and surface reactivity (reflected by ICE) provide secondary but essential contributions. Other parameters, including mesoporosity and tap density, remain implicitly coupled with SSA within the present dataset, as higher surface area materials are generally associated with increased porosity and reduced structural compactness. This intrinsic coupling explains their indirect influence on the observed trade-offs [47], [61].

Overall, the proposed framework demonstrates that interpretable machine learning, when applied to carefully curated experimental datasets, can provide physically meaningful insights into complex structure–property relationships. As larger and more diverse datasets become available, this approach can be further refined to establish more general and quantitatively robust design

rules for carbon-based anodes in sodium-ion and related battery systems [23], [28], [61].

IV. CONCLUSION

This work demonstrates that interpretable machine learning, when integrated with carefully controlled experimental characterization, can uncover physically meaningful structure–property relationships in carbon-based anodes beyond conventional trial-and-error approaches. The analysis consistently identifies specific surface area (SSA) as the primary descriptor governing electrochemical performance, with additional contributions from structural ordering (interlayer spacing, $d_{\{002\}}$) and surface reactivity (reflected by ICE), revealing a constrained multi-objective trade-off between capacity, rate capability, and efficiency.

By combining experimental observations with model interpretability, this study establishes a practical framework for translating complex descriptor interactions into actionable design guidelines. The results show that optimal performance does not arise from maximizing a single parameter, but from balancing surface development, electrochemical reactivity, and structural compactness within a defined design window.

Within the explored parameter space, an optimal regime is identified at moderate SSA ($\sim 400\text{--}650\text{ m}^2\text{ g}^{-1}$), where reversible capacities above $\sim 270\text{--}300\text{ mAh g}^{-1}$ can be achieved while maintaining ICE above $\sim 80\%$ and acceptable cycling stability. This finding highlights that sodium storage in disordered carbon systems is governed by a non-linear and physically bounded design space.

We acknowledge that these conclusions are derived from a compact dataset ($n = 12$), which limits statistical generalization and confines the validity of the design rules to the investigated material domain. Future work will focus on expanding the dataset through integration of additional experimental and literature data, incorporating advanced descriptors (e.g., heteroatom speciation and pore connectivity), and extending the framework toward multi-objective optimization and full-cell validation.

Overall, this study demonstrates that interpretable machine learning provides a robust and scalable pathway for linking synthesis, structure, and performance, enabling rational and mechanism-driven design of carbon-based anodes for sodium-ion and related rechargeable battery systems.^[8]

ACKNOWLEDGMENTS

The author acknowledges Universitas Muhammadiyah PKU Surakarta for providing research facilities and institutional support. The author also thanks the Electrical Laboratory, Faculty of Science and Technology, for technical assistance in materials characterization and electrochemical testing.

FUNDING

This research received no external funding.

ACKNOWLEDGMENTS

The authors acknowledge Universitas Muhammadiyah PKU Surakarta for institutional support.

REFERENCES

- [1] S. W. Kim, D. H. Seo, X. Ma, G. Ceder, and K. Kang,^[12] “Electrode materials for rechargeable sodium-ion batteries: Potential alternatives to current lithium-ion batteries,” Jul. 2012. doi: 10.1002/aenm.201200026.
- [2] K. Chayambuka, G. Mulder, D. L. Danilov, and R. H. L. Notten, “Sodium-Ion Battery Materials and Electrochemical Properties Reviewed,” *Adv. Energy Mater.*, vol. 8, no. 16, p. 1800079, Jun. 2018, doi: 10.1002/AENM.201800079;REQUESTEDJOURNAL:JOURNAL:16146840;WGROU:STRING:PUBLICATION.^[13]
- [3] Y. Zhou and Z. Shen, “From Sodium Storage Mechanism to Design of High-Capacity Carbon-Based Anode: A Review,” *Materials* 2025, Vol. 18, Page 2248, vol. 18, no. 10, p. 2248, May 2025, doi: 10.3390/MA18102248.^[14]
- [4] X. Dou *et al.*, “Hard carbons for sodium-ion batteries: Structure, analysis, sustainability, and electrochemistry,” *Materials Today*, vol. 23, pp. 87–104, Mar. 2019, doi: 10.1016/J.MATTOD.2018.12.040.^[15]
- [5] D. Chen *et al.*, “Hard carbon for sodium storage: mechanism and optimization strategies toward commercialization,” *Energy Environ. Sci.*, vol. 14, no. 4, pp. 2244–2262, Apr. 2021, doi: 10.1039/D0EE03916K.^[16]
- [6] M. Thompson, Q. Xia, Z. Hu, and X. S. Zhao, “A review on biomass-derived hard carbon materials for sodium-ion batteries,” *Mater. Adv.*, vol. 2, no. 18, pp. 5881–5905, Sep. 2021, doi: 10.1039/D1MA00315A.
- A. Funke and F. Ziegler, “Hydrothermal carbonization of biomass: A summary and discussion of chemical mechanisms for process engineering,” *Biofuels, Bioproducts and Biorefining*, vol. 4, no. 2, pp. 160–177, Mar. 2010, doi: 10.1002/BBB.198;PAGE:STRING:ARTICLE/CHAPTER.
- B. Zhong *et al.*, “Biomass-Derived Hard Carbon for Sodium-Ion Batteries: Basic Research and Industrial Application,” *ACS Nano*, vol. 18, no. 26, pp. 16468–16488, Jul. 2024, doi: 10.1021/ACS.NANO.4C03484.
- N. Kitchamsetti, K. H. Kim, H. S. Han, and S. Mhin, “Biomass-Derived Hard Carbon Anodes for Sodium-Ion Batteries: Recent Advances in Synthesis Strategies,” *Nanomaterials* 2025, Vol. 15, Page 1554, vol. 15, no. 20, p. 1554, Oct. 2025, doi: 10.3390/NANO15201554.
- S. Nour Eddine *et al.*, “From Biomass to Energy Storage: A Review on Lignocellulosic Biomass-Derived Hard Carbon Anodes for Sodium-Ion Batteries,” *Energy & Fuels*, vol. 39, no. 36, pp. 17112–17154, Sep. 2025, doi: 10.1021/ACS.ENERGYFUELS.5C02262.
- Z. Li, Q. Lu, Z. Xu, Y. Wang, and H. L. Wang, “Development of Biomass-Derived Hard Carbon for Sodium-Ion Batteries: Optimizing Microstructure and Electrochemical Performance,” *Electroanalysis*, vol. 37, no. 1, p. e202400369, Jan. 2025, doi: 10.1002/ELAN.202400369.
- S. Gan *et al.*, “Comprehensive Understanding of Closed Pores in Hard Carbon Anode for High-Energy Sodium-Ion Batteries,” *Nano-Micro Letters* 2025 17:1, vol. 17, no. 1, pp. 325–, Jul. 2025, doi: 10.1007/S40820-025-01833-X.
- Y. Guo *et al.*, “Advancing hard carbon anode for sodium-ion batteries: Mechanisms and optimization strategies,” *Nano Research Energy*, vol. 4, no. 2, p. e9120165, Jun. 2025, doi: 10.26599/NRE.2025.9120165.
- J. Wang *et al.*, “Hard carbon anodes for advanced sodium ion batteries: A review on sodium storage mechanism and strategies to improve the initial Coulombic efficiency,” *Chemical Engineering Journal*, vol. 511, p. 161953, May 2025, doi: 10.1016/J.CEJ.2025.161953.
- T. Gao and Y. Ding, “From biomass to batteries: Cutting-edge advances in hard carbon anodes for sustainable sodium-ion energy storage,” *Inorg. Chem. Commun.*, vol. 180, p. 114925, Oct. 2025, doi: 10.1016/J.INOCHE.2025.114925.
- J. Wang and S. Kaskel, “KOH activation of carbon-based materials for energy storage,” *J. Mater. Chem.*, vol. 22, no. 45, pp. 23710–23725, Oct. 2012, doi: 10.1039/C2JM34066F.
- P. Liu, Y. Li, Y. S. Hu, H. Li, L. Chen, and X. Huang, “A waste biomass derived hard carbon as a high-performance anode material for sodium-ion batteries,” *J. Mater. Chem. A Mater.*, vol. 4, no. 34, pp. 13046–13052, Aug. 2016, doi: 10.1039/C6TA04877C.

- [18] K. L. Hong *et al.*, “Biomass derived hard carbon used as a high performance anode material for sodium ion batteries,” *J. Mater. Chem. A Mater.*, vol. 2, no. 32, pp. 12733–12738, Jul. 2014, doi: 10.1039/C4TA02068E.
- [19] F. Wu, M. Zhang, Y. Bai, X. Wang, R. Dong, and C. Wu, “Lotus Seedpod-Derived Hard Carbon with Hierarchical Porous Structure as Stable Anode for Sodium-Ion Batteries,” *ACS Appl. Mater. Interfaces*, vol. 11, no. 13, pp. 12554–12561, Apr. 2019, doi: 10.1021/ACSAMI.9B01419.
- [20] Y. Zhang *et al.*, “Honeycomb-like Hard Carbon Derived from Pine Pollen as High-Performance Anode Material for Sodium-Ion Batteries,” *ACS Appl. Mater. Interfaces*, vol. 10, no. 49, pp. 42796–42803, Dec. 2018, doi: 10.1021/ACSAMI.8B13160.
- [21] Y. Tang *et al.*, “Electrochemical Behavior of the Biomass Hard Carbon Derived from Waste Corn cob as a Sodium-Ion Battery Anode,” *Energy & Fuels*, vol. 38, no. 8, pp. 7389–7398, Apr. 2024, doi: 10.1021/ACS.ENERGYFUELS.4C00564.
- [22] X. Liu *et al.*, “A Fast Forward Prediction Framework for Energy Materials Design Based on Machine Learning Methods,” *Energy Material Advances*, vol. 2024, Dec. 2024, doi: 10.34133/ENERGYMATADV.0131;JOURNAL:JOURNAL:ENERGYMATADV;WGROU:STRING:PUBLICATION.
- [23] C. Ling, “A review of the recent progress in battery informatics,” *npj Computational Materials* 2022 8:1, vol. 8, no. 1, pp. 33-, Feb. 2022, doi: 10.1038/s41524-022-00713-x.
- [24] A. Yadav, V. Kumar, S. Singh, and A. Kumar Mishra, “Machine Learning in Lithium-Ion Battery: Applications, Challenges, and Future Trends,” *SN Computer Science* 2024 5:6, vol. 5, no. 6, pp. 717-, Jul. 2024, doi: 10.1007/S42979-024-03046-2.
- [25] L. Li *et al.*, “Predicting Cycle Life for Lithium-Ion Batteries with Ternary Cathode Materials Using Data-Driven Machine Learning,” *ACS Omega*, vol. 10, pp. 52001–52009, Oct. 2025, doi: 10.1021/ACSOMEGA.5C09364.
- [26] N. Noura, L. Boulon, and S. Jemeï, “A Review of Battery State of Health Estimation Methods: Hybrid Electric Vehicle Challenges,” *World Electric Vehicle Journal* 2020, Vol. 11, Page 66, vol. 11, no. 4, p. 66, Oct. 2020, doi: 10.3390/WEVJ11040066.
- [27] L. Mbagaya, K. Reddy, and A. Botes, “Machine Learning Techniques for Battery State of Health Prediction: A Comparative Review,” *World Electric Vehicle Journal* 2025, Vol. 16, Page 594, vol. 16, no. 11, p. 594, Oct. 2025, doi: 10.3390/WEVJ16110594.
- [28] M. Bao *et al.*, “Interpretable machine learning prediction for li-ion battery’s state of health based on electrochemical impedance spectroscopy and temporal features,” *Electrochim. Acta*, vol. 494, p. 144449, Aug. 2024, doi: 10.1016/J.ELECTACTA.2024.144449.
- [29] X. Li *et al.*, “Structural design of biomass-derived hard carbon anode materials for superior sodium storage via increasing crystalline cellulose and closing the pores,” *J. Mater. Chem. A Mater.*, vol. 12, no. 32, pp. 21176–21189, Aug. 2024, doi: 10.1039/D4TA03284E.
- R. R. Gaddam, A. H. Farokh Niaei, M. Hankel, D. J. Searles, N. A. Kumar, and X. S. Zhao, “Capacitance-enhanced sodium-ion storage in nitrogen-rich hard carbon,” *J. Mater. Chem. A Mater.*, vol. 5, no. 42, pp. 22186–22192, Oct. 2017, doi: 10.1039/C7TA06754B.
- C. Chen, Y. Huang, Z. Meng, Z. Xu, P. Liu, and T. Li, “Multi-heteroatom doped porous carbon derived from insect feces for capacitance-enhanced sodium-ion storage,” *Journal of Energy Chemistry*, vol. 54, pp. 482–492, Mar. 2021, doi: 10.1016/J.JECHEM.2020.06.025.
- J. Kuai *et al.*, “Comparison and optimization of biomass-derived hard carbon as anode materials for sodium-ion batteries,” *Chem. Phys. Lett.*, vol. 842, p. 141214, May 2024, doi: 10.1016/J.CPLETT.2024.141214.
- R. Dong *et al.*, “Tailoring Defects in Hard Carbon Anode towards Enhanced Na Storage Performance,” *Energy Material Advances*, vol. 2022, 2022, doi: 10.34133/2022/9896218.
- N. Voronina and S. T. Myung, “Recent Advances in Electrode Materials with Anion Redox Chemistry for Sodium-Ion Batteries,” *Energy Material Advances*, vol. 2021, Jul. 2021, doi: 10.34133/2021/9819521.
- X. Shen *et al.*, “Advanced Electrode Materials in Lithium Batteries: Retrospect and Prospect,” *Energy Material Advances*, vol. 2021, 2021, doi: 10.34133/2021/1205324.
- F. Wu, J. Liu, I. 2019, J. B. Goodenough, M. S. Whittingham, and A. Yoshino, “Energy Material Advances: From Fundamental Discoveries to Practical Applications,” *Energy Material Advances*, vol. 2020, Dec. 2020, doi: 10.34133/2020/7981289.
- Z. Wang *et al.*, “High-Energy Na₄MnCr(PO₄)₃@C Cathode for Solid-State Sodium Metal Batteries,” *Energy Material Advances*, vol. 4, Jun. 2023, doi: 10.34133/ENERGYMATADV.0036.
- J. Yue *et al.*, “Initiating High-Voltage Multielectron Reactions in NASICON Cathodes for Aqueous Zinc/Sodium Batteries,” *Energy Material Advances*, vol. 4, Aug. 2023, doi: 10.34133/ENERGYMATADV.0050;WEBSITE:WEBSITE:SPJ-SITE;REQUESTEDJOURNAL:JOURNAL:ENERGYMATADV;JOURNAL:JOURNAL:ENERGYMATADV;ISSUE:ISSUE:DOI.
- Y. Dong *et al.*, “Challenges and Strategies of Fast-Charging Li-Ion Batteries with a Focus on Li Plating,” *Energy Material Advances*, vol. 5, Aug. 2024, doi: 10.34133/ENERGYMATADV.0113.
- X. Zhang, S. Chen, J. Zhu, and Y. Gao, “A Critical Review of Thermal Runaway Prediction and Early-Warning Methods for Lithium-Ion Batteries,” *Energy Material Advances*, vol. 4, Jan. 2023, doi: 10.34133/ENERGYMATADV.0008;PAGE:STRING:ARTICLE/CHAPTER.
- F. Wu *et al.*, “Removing the Intrinsic NiO Phase and Residual Lithium for High-Performance Nickel-Rich

- Materials,” *Energy Material Advances*, vol. 4, Jan. 2023, doi: 10.34133/ENERGYMATADV.0007;CTYPE:STRING:G:JOURNAL.
- [42] G. Lee, J. Kim, and C. Lee, “State-of-health estimation of Li-ion batteries in the early phases of qualification tests: An interpretable machine learning approach,” *Expert Syst. Appl.*, vol. 197, p. 116817, Jul. 2022, doi: 10.1016/J.ESWA.2022.116817.
- [43] M. A. Hannan *et al.*, “Toward Enhanced State of Charge Estimation of Lithium-ion Batteries Using Optimized Machine Learning Techniques,” *Scientific Reports* 2020 10:1, vol. 10, no. 1, pp. 4687–, Mar. 2020, doi: 10.1038/s41598-020-61464-7.
- [44] S. M. Lundberg and S. I. Lee, “A Unified Approach to Interpreting Model Predictions,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-December, pp. 4766–4775, May 2017, Accessed: Nov. 29, 2025. [Online] Available: <https://arxiv.org/pdf/1705.07874>
- [45] K. Nikgoftar, A. K. Madikere Raghunatha Reddy, M. V. Reddy, and K. Zaghib, “Carbonaceous Materials as Anodes for Lithium-Ion and Sodium-Ion Batteries,” *Batteries* 2025, Vol. 11, Page 123, vol. 11, no. 4, p. 123, Mar. 2025, doi: 10.3390/BATTERIES11040123.
- [46] D. Alvira, D. Antorán, and J. J. Manyà, “Plant-derived hard carbon as anode for sodium-ion batteries: A comprehensive review to guide interdisciplinary research,” *Chemical Engineering Journal*, vol. 447, p. 137468, Nov. 2022, doi: 10.1016/J.CEJ.2022.137468.
- [47] X. Li *et al.*, “Progress in hard carbons for sodium-ion batteries: Microstructure, sodium storage mechanism and initial Coulombic efficiency,” *J. Energy Storage*, vol. 98, Sep. 2024, doi: 10.1016/J.EST.2024.112989.
- [48] N. Sun, J. Qiu, and B. Xu, “Understanding of Sodium Storage Mechanism in Hard Carbons: Ongoing Development under Debate,” *Adv. Energy Mater.*, vol. 12, no. 27, p. 2200715, Jul. 2022, doi: 10.1002/AENM.202200715.
- [49] H. D. Asfaw, C. W. Tai, M. Valvo, and R. Younesi, “Facile synthesis of hard carbon microspheres from polyphenols for sodium-ion batteries: insight into local structure and interfacial kinetics,” *Mater. Today Energy*, vol. 18, p. 100505, Dec. 2020, doi: 10.1016/J.MTENER.2020.100505.
- [50] R. Alcántara, P. Lavela, G. F. Ortiz, and J. L. Tirado, “Carbon Microspheres Obtained from Resorcinol-Formaldehyde as High-Capacity Electrodes for Sodium-Ion Batteries,” *Electrochemical and Solid-State Letters*, vol. 8, no. 4, p. A222, Mar. 2005, doi: 10.1149/1.1870612.
- [51] X. Dou *et al.*, “Pectin, Hemicellulose, or Lignin? Impact of the Biowaste Source on the Performance of Hard Carbons for Sodium-Ion Batteries,” *ChemSusChem*, vol. 10, no. 12, pp. 2668–2676, Jun. 2017, doi: 10.1002/CSSC.201700628;PAGE:STRING:ARTICLE/CHAPTER.
- [52] Y. Feng *et al.*, “Chemical-enzymatic fractionation to unlock the potential of biomass-derived carbon materials for sodium ion batteries,” *J. Mater. Chem. A Mater.*, vol. 7, no. 47, pp. 26954–26965, Dec. 2019, doi: 10.1039/C9TA09124F.
- Z. Li *et al.*, “High Capacity of Hard Carbon Anode in Na-Ion Batteries Unlocked by POx Doping,” *ACS Energy Lett.*, vol. 1, no. 2, pp. 395–401, Aug. 2016, doi: 10.1021/ACSENERGYLETT.6B00172.
- L. Qie *et al.*, “Sulfur-Doped Carbon with Enlarged Interlayer Distance as a High-Performance Anode Material for Sodium-Ion Batteries,” *Adv. Sci. (Weinh.)*, vol. 2, no. 12, Dec. 2015, doi: 10.1002/ADVS.201500195.
- B. Yan, C. Han, Y. Dai, M. Li, Z. Wu, and X. Gao, “Biomass derived hard carbon materials for sodium ion battery anodes: Exploring the influence of carbon source on structure and sodium storage performance,” *Fuel*, vol. 371, p. 132141, Sep. 2024, doi: 10.1016/J.FUEL.2024.132141.
- A. Kumar *et al.*, “Biomass-based carbon material for next-generation sodium-ion batteries: insights and SWOT evaluation,” *Environ. Res.*, vol. 279, p. 121854, Aug. 2025, doi: 10.1016/J.ENVRES.2025.121854.
- H. Moon *et al.*, “Bio-Waste-Derived Hard Carbon Anodes Through a Sustainable and Cost-Effective Synthesis Process for Sodium-Ion Batteries,” *ChemSusChem*, vol. 16, no. 1, Jan. 2023, doi: 10.1002/CSSC.202201713.
- C. Nita, B. Zhang, J. Dentzer, and C. Matei Ghimbeu, “Hard carbon derived from coconut shells, walnut shells, and corn silk biomass waste exhibiting high capacity for Na-ion batteries,” *Journal of Energy Chemistry*, vol. 58, pp. 207–218, Jul. 2021, doi: 10.1016/J.JEACHEM.2020.08.065.
- N. T. Aristote, Z. Song, W. Deng, H. Hou, G. Zou, and X. Ji, “Effect of double and triple-doping of sulfur, nitrogen and phosphorus on the initial coulombic efficiency and rate performance of the biomass derived hard carbon as anode for sodium-ion batteries,” *J. Power Sources*, vol. 558, p. 232517, Feb. 2023, doi: 10.1016/J.JPOWSOUR.2022.232517.
- Y. Chen *et al.*, “Sustainable and scalable fabrication of high-performance hard carbon anode for Na-ion battery,” *J. Power Sources*, vol. 557, p. 232534, Feb. 2023, doi: 10.1016/J.JPOWSOUR.2022.232534.
- T. Qi, X. Zhang, K. Xiong, H. Yang, S. Zhang, and H. Chen, “Prediction of discharge performance of hard carbon materials for high-performance sodium-ion batteries based on machine learning,” *J. Mater. Chem. A Mater.*, vol. 13, no. 23, pp. 17748–17771, Jun. 2025, doi: 10.1039/D5TA01528F.
- S. Y. Owusu, “Machine learning prediction of the reversible capacities of a biomass-derived hard carbon anode for sodium-ion batteries,” *RSC Sustainability*, vol. 3, no. 7, pp. 3133–3143, Jul. 2025, doi: 10.1039/D5SU00360A.