ChemPlusChem

Research Article
doi.org/10.1002/cplu.202500342

Chemistry Europe
European Chemical Societies Publishing

www.chempluschem.org

Machine Learning—Guided Design of Biomass-Based Porous Carbon for Aqueous Symmetric Supercapacitors

Manickam Minakshi,* Apsana Sharma, Ferdous Sohel, Almantas Pivrikas, Pragati A. Shinde, Katsuhiko Ariga, and Lok Kumar Shrestha*

Biomass-derived porous carbon electrodes have attracted significant attention for high-performance supercapacitor applications due to their sustainability, cost-effectiveness, and tunable porosity. To accelerate the design and evaluation of these materials, it is essential to develop accurate and efficient strategies for optimizing their physicochemical and electrochemical properties. Herein, a machine learning (ML) approach is employed to analyze experimental data from previously reported sources, enabling the prediction of specific capacitance (F g^{-1}) based on various material characteristics and processing conditions. The trained ML model evaluates the influence of factors such as biomass type, electrolyte, activating agent, and key synthesis parameters, including activation and carbonization temperatures and durations, on supercapacitor performance. Despite growing interest, comprehensive studies that correlate these variables with performance metrics remain limited. This work addresses this gap by using ML algorithms to uncover the interrelationships between biomass-derived carbon properties, synthesis conditions, and specific capacitance. Herein, it is demonstrated that an optimal combination of a carbonized honeydew peel to H_3PO_4 ratio of 1:4 and an activation temperature of 500 °C yields a highly porous carbon material. When used in a symmetric device with 1 M H_2SO_4 electrolyte, this material, rich in oxygen and phosphorus species, achieves a high specific capacitance of 611 F g^{-1} at a current density of 1.3 A g^{-1}. Correlation analysis reveals a strong synergy between surface area and pore volume (correlation coefficient = 0.8473), and the ML-predicted capacitance closely aligns with experimental results. This ML-assisted framework offers valuable insights into the critical physicochemical and electrochemical parameters that govern supercapacitor performance, providing a powerful tool for the rational design of next-generation energy storage materials.

1. Introduction

Rechargeable battery technologies are well-established and widely utilized in consumer electronics and electric vehicles. However, they face several limitations, including substantial weight, large volume, high internal resistance, low power density, and limited transient response. In contrast, supercapacitors, also known as ultracapacitors or electrochemical double-layer capacitors (EDLCs), have emerged as promising alternatives for energy storage, owing to their rapid charge–discharge capability, high power density, low weight, compact size, and low internal resistance.[1]

Although supercapacitors have been known since the 1950s, their potential has only recently been fully recognized.[2] Among various types, aqueous supercapacitors are particularly attractive due to their low cost, environmental friendliness, and excellent cycling stability, making them ideal for applications such as regenerative braking systems in electric vehicles. The performance of supercapacitors is largely governed by the properties of the electrode materials, which directly influence capacitance, energy density, and overall device efficiency. Designing and optimizing electrode materials is, therefore, critical to advancing supercapacitor technology.[3] Carbon-based materials are especially well-suited for this purpose, offering high electrical

M. Minakshi, A. Sharma, F. Sohel, A. Pivrikas
College of Science
Technology, Engineering & Mathematics
Murdoch University
Murdoch 6150, Western Australia, Australia
E-mail: minakshi@murdoch.edu.au

M. Minakshi, P. A. Shinde, K. Ariga, L. K. Shrestha
Research Center for Materials Nanoarchitectonics (MANA)
National Institute for Materials Science (NIMS)
Tsukuba 305 0044, Japan
E-mail: shrestha.lokkumar@nims.go.jp

K. Ariga
Department of Advanced Materials Science
Graduate School of Frontier Sciences
The University of Tokyo
5-1-5 Kashiwanoha, Kashiwa 277–8561, Chiba, Japan

L. K. Shrestha
Department of Materials Science
Institute of Pure and Applied Sciences
University of Tsukuba
1–1 Tennodai, Tsukuba 305–8573, Ibaraki, Japan

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cplu.202500342

conductivity, chemical stability, affordability, and long cycle life.[4] In particular, porous carbon materials derived from biomass have gained attention due to their sustainability and tunable textural properties, which are essential for enhancing ion transport and charge storage in aqueous electrolytes.

Biochar, produced through the thermochemical conversion of biomass, provides a sustainable and eco-friendly platform for synthesizing a wide range of functional carbon materials, including porous carbon, heteroatom-doped biochar, carbon nanotubes, graphene, and carbon quantum dots, for advanced applications.[5] The use of biomass-derived carbon in energy storage has gained significant attention due to its cost-effectiveness, ease of production, and environmental sustainability. Meso- and microporous carbons can be readily obtained from diverse agricultural and biological sources such as plants, fruits, microorganisms, and animal residues.[6] Activated biochar-based electrode materials, prepared via pyrolysis, hydrothermal treatment, chemical activation, and heteroatom doping, have emerged as promising candidates for supercapacitor applications.[7] These materials exhibit favorable physical and chemical properties, including well-developed pore structures, abundant surface functional groups, and excellent energy storage capabilities.[8] In contrast, biochar produced through direct carbonization often suffers from limited electrochemical performance due to its underdeveloped porosity, low graphitization, and insufficient surface functionality.[9]

The energy storage performance of activated biochar is influenced by multiple factors, including the intrinsic properties of the biomass precursor, activation conditions, pore architecture, and testing parameters. The complexity of these interdependent variables makes it challenging to identify optimal synthesis pathways through experimental methods alone.[10] For instance, Jia et al.[11] conducted 20 carbonization experiments on different parts of corn straw, revealing that moderate carbonization temperatures and low ash content positively affect biochar's energy storage performance. Similarly, Koutcheiko et al.[12] demonstrated that optimizing activation conditions can enhance capacitance, but the process remains labor-intensive and lacks standardized guidelines.

To address these challenges, data-driven approaches such as machine learning (ML) offer powerful tools for modeling and predicting the relationships between synthesis parameters and electrochemical performance. ML techniques can simulate complex interactions between input variables and output metrics, enabling efficient optimization of material properties.[13] Leng et al.[14] applied random forest (RF) and gradient boosting regression models to predict biochar yield and surface area, identifying pyrolysis temperature as a key factor. Yang et al.[15] used ML to correlate activation conditions with specific capacitance, highlighting the importance of surface area and activator dosage. Despite these advances, comprehensive optimization of the full synthesis chain, from biomass selection to activation and doping, remains elusive. Repeated experimentation to achieve desired material properties is time-consuming and inefficient. Therefore, further investigation is needed to understand how different biomass types and processing conditions influence the electrochemical behavior of activated biochar.[16]

In this study, we perform a feature importance analysis to identify the most influential parameters, such as biomass type, activating agent ratio, and activation temperature, on specific capacitance. We extend our ML model to predict the capacitance of honeydew peel-derived porous carbon in acid-aqueous electrolytes, building on our previous work,[17] which demonstrated a significant improvement in specific capacitance from 185 to 485 F g^{-1} in single-electrode configurations. A comparative analysis was conducted using four ML models, RF, Support Vector Regression (SVR), Artificial Neural Network (ANN), and Multi-Layer Perceptron (MLP), to predict the capacitance of biomass-derived carbon supercapacitors. Eleven features were categorized into material properties (biomass type, surface area, pore size), thermal treatment parameters (activation/carbonization temperatures and durations), and electrochemical conditions (activating agent, electrolyte, voltage window, current density). The capacitive contributions from pores and heteroatoms were evaluated based on activator dosage and synthesis conditions.

Overall, ML techniques were employed to predict and analyze the influence of biomass carbon properties and experimental parameters on supercapacitor performance. A comprehensive dataset was compiled and validated using honeydew peel-derived porous carbon. Among the models tested, SVR achieved the highest coefficient of determination ($R^2 = 0.41398$) with a mean absolute error (MAE = 0.11359). Both SVR and RF models demonstrated strong predictive capabilities, achieving determination coefficients (R^2) of 0.41396 and 0.35689, MAEs of 0.11359 and 0.11054, and root mean square errors (RMSE) of 0.14897 and 0.13734, respectively. The ML classification algorithms highlighted key contributors to supercapacitor performance, with surface area and pore volume exhibiting strong correlation ratios of 0.8473 each. The optimal combination of honeydew peel and activator ratio, along with suitable synthesis temperature, resulted in functionalized carbon surfaces enriched with phosphorus and oxygen, enhancing wettability and pseudocapacitance for high-performance aqueous supercapacitors.

2. Methodology

2.1. ML Approach

2.1.1. Dataset Description

To evaluate and predict the specific capacitance of biomass-derived carbon-based supercapacitors, ML techniques were employed to analyze the influence of various material and process parameters. Feature importance scoring, commonly used in regression and classification models, was applied to rank the contribution of each input variable to the target output (specific capacitance). This approach enables researchers to identify the most influential features and reduce the dimensionality of the dataset without compromising model performance. To ensure model robustness and prevent overfitting, cross-validation was

implemented. This technique partitions the dataset into multiple folds, training the model on a subset while validating it on the remaining data. The process is repeated across different folds, providing a more reliable estimate of model generalizability to unseen data.[18]

2.1.2. Data Collection and Preprocessing

The dataset was curated from previously reported literature (see Table S1, Supporting Information), with a focus on maintaining consistency and reducing complexity. To ensure homogeneity, only aqueous electrolytes such as KOH and H_2SO_4 were included, while organic electrolytes and ionic liquids (e.g., $TEABF_4$) were excluded.[19] Since specific capacitance varies with current density, normalization was applied to account for differences across studies (**Table 1**).

2.1.3. Data Segregation

Variables were categorized into continuous and categorical types to guide appropriate preprocessing and model selection. Continuous variables (e.g., surface area, pore size) were normalized or scaled to ensure uniform learning, while categorical variables (e.g., biomass type, electrolyte) were encoded using one-hot or label encoding techniques.[20] The dataset was further structured into independent variables (features) and a dependent variable (specific capacitance), as summarized in **Table 2**. This distinction is critical for supervised learning, where the model learns to map input features to the target output.[21] The eleven features used in the model were grouped into three categories: 1) **Material Properties**: Biomass type, surface area, and pore size; 2) **Thermal Treatment Parameters**: Activation temperature, carbonization temperature, and activation time, carbonization time; 3) **Electrochemical Parameters**: Activating agent, electrolyte type, voltage window, and current density.

2.1.4. Data Processing

In this study, MATLAB software was employed for all stages of data handling, including preprocessing, model training, evaluation, and feature importance analysis aimed at predicting specific capacitance. Data handling and preprocessing were conducted to import and structure the dataset efficiently. For regression modeling, the Support Vector Machine algorithm with a Gaussian kernel was implemented using MATLAB's fitcsvm function, enabling nonlinear mapping of input features to the target variable. To enhance interpretability and reduce dataset complexity, Principal Component Analysis (PCA) was applied using MATLAB's PCA function. This technique facilitated improved visualization of feature distributions and relationships. Various plots, including scatter plots, regression surfaces, and feature correlation matrices, were generated using MATLAB's built-in plotting tools to support exploratory data analysis and model interpretation.[22] Feature importance was assessed using permutation feature importance, which quantifies the increase in prediction error when individual features are randomly permuted. This method provides insights into the relative contribution of each feature to model accuracy, helping identify the most influential parameters affecting specific capacitance.[23,24] A schematic overview of the ML workflow is presented in **Figure 1**.

Table 1. Variables are segregated into categorical and continuous variables.

Categorical variable	Continuous variable
Biomass material;	Activation temperature;
Activating agent; and	Activation time;
Electrolyte	Carbonization temperature;
	Carbonization time;
	Pore volume;
	Surface area;
	Current density;
	Voltage window, and
	Specific capacitance

Table 2. Variables are segregated into independent and dependent variables.

Independent Variable	Dependent Variable
Biomass material;	Capacitance
Activating agent; Electrolyte;	
Activation temperature;	
Activation time;	
Carbonization temperature;	
Carbonization time;	
Pore volume;	
Surface area;	
Current density; and	
Voltage window	

2.1.5. Data Cleaning

In this study, data normalization was employed as a critical processing step to scale numerical features to a consistent range, typically between 0 and 1. This ensures that variables with differing

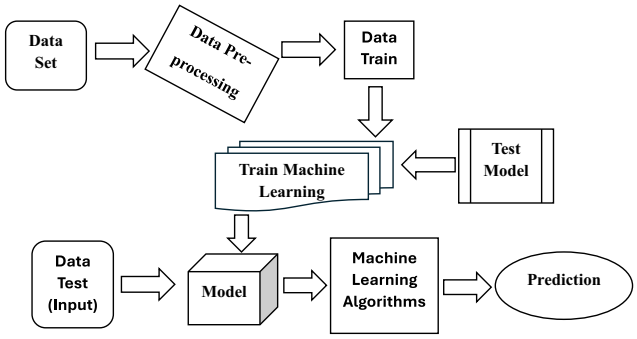

Figure 1. Schematic representation of the ML workflow involved in model prediction.

units or magnitudes do not disproportionately influence the learning process of the ML model. Normalization enhances both the performance and stability of ML algorithms by enabling uniform treatment of all input features.[25] Additionally, data preprocessing was undertaken to prepare raw data for analysis, which involved cleaning (handling missing values, removing noise), transforming (normalizing and encoding categorical variables), and organizing it for optimal model training. This step ensured that the data was clean, consistent, and adequately structured for effective learning by the ML model.[26] Continuous variables such as surface area, pore volume, and temperature were normalized to maintain a uniform scale across the dataset. Categorical variables were encoded using appropriate techniques to ensure compatibility with the ML models.

To prioritize features with the highest predictive potential, correlation analysis was performed. This helped identify variables consistently reported as significant in influencing specific capacitance, guiding feature selection and model refinement.[26]

2.2. Physical Characterization of Heteroatom-Doped Honeydew Peel

The structural properties of the synthesized honeydew peel-derived activated carbon (HDP-AC), prepared at a synthesis temperature of 500 °C, were investigated using X-ray diffraction (Rigaku, Japan) using Cu Kα radiation. To assess the nature of carbon bonding and surface structure, Raman spectroscopy was performed using an NRS-3100 (JASCO, Tokyo, Japan) instrument with a 532.09 nm neon laser excitation wavelength. Each sample was analyzed at a depth of 3 μm across at least three different surface locations to ensure representative data. The surface morphology and elemental composition of HDP-AC were examined using Field Emission Scanning Electron Microscopy (FE-SEM) (TESCAN CLARA). For high-resolution imaging of internal structures, Transmission Electron Microscopy (TEM) was conducted using a JEOL 2200FS TEM operated at 200 kV. TEM specimens were prepared by dispersing HDP-AC powder in ethanol, followed by grinding in an agate mortar and pipetting onto a holey carbon film supported on a copper mesh grid.

2.3. Electrochemical Characterization of Heteroatom-Doped Honeydew Peel

Electrochemical characterization of the HDP-AC was performed using both three-electrode and symmetric two-electrode configurations. The working electrode was fabricated by mixing HDP-AC (75 wt%), carbon black (15 wt%), and polyvinylidene difluoride (PVDF, 10 wt%) in 0.4 mL of N-Methyl-2-pyrrolidone (NMP) to form a homogeneous slurry. This slurry was coated onto a graphite sheet with an active area of 1 cm^2, and the mass of active material loaded per electrode was ≈2 mg. In three-electrode tests, platinum wire of 10 cm in length and 1 mm in diameter dimension and mercury–mercuric oxide (Hg/HgO) served as the counter and reference electrode, respectively. In single-electrode tests, HDP-AC served as the working electrode. The cyclic voltammetry (CV) and galvanostatic charge–discharge studies of the composites were carried out using SP-150, Bio-Logic Science instruments in 1 M H_2SO_4 electrolyte at room temperature. For HDP-AC, the working electrode was cycled between 0 and 1.1 V. The frequency range for electrochemical impedance spectroscopy (EIS) was 10 40 MHz with a 5 mV bias voltage.

The specific capacitance of the symmetric supercapacitor device was computed from the following (Equation 1).

$$C_s = \frac{i \times \Delta t}{\Delta v \times m} (F g^{-1}) \quad (1)$$

where C_s is the specific capacitance in ($F\,g^{-1}$), i is the current in (mA), Δt is the discharging time (s), m is the mass of the active material loaded (mg), and ΔV is the potential window (V). The energy and power density of the device were calculated using the formulas in (Equations 2) and (3).

$$E = \frac{1}{8} C_s \frac{\Delta v^2}{3.6} (Wh\,kg^{-1}) \quad (2)$$

$$P = \frac{E \times 3600}{t} (W\,kg^{-1}) \quad (3)$$

where E and P denote the energy density (Wh Kg^{-1}) and power density (W Kg^{-1}) of the device. For the symmetric device, two identical HDP-AC electrodes (each 2 mg) were assembled in a two-electrode configuration using 1 M H_2SO_4 as the electrolyte.

3. Results

3.1. ML

Multiple ML models were evaluated using three key performance metrics: RMSE, MAE, and the Coefficient of Determination (R^2). These metrics provided a robust framework for assessing the predictive accuracy of each model and its alignment with experimental values. The training dataset comprised electrochemical performance data of heteroatom-doped HDP-AC electrodes, synthesized with varying activating agent ratios and tested in 1 M H_2SO_4 electrolyte. This dataset enabled the models to learn the relationships between synthesis parameters and specific capacitance, facilitating reliable predictions and feature importance analysis. The proposed ML model was trained on a dataset comprising 166 data points, as detailed in Table S1, Supporting Information. To ensure robust model training and validation, the dataset was partitioned into training, validation, and test sets, each representing approximately 20% of the total data. RMSE is assessed, with lower values indicating better model fit. R-squared (R^2) values are analyzed to determine how well the input features explain the variance in the output variable, with higher values indicating stronger predictive performance.[27] MAE is also examined, measuring the average absolute difference between predicted and actual values, providing insight into overall prediction accuracy. To further ensure generalizability, fourfold cross-validation was applied. This technique tests the model's

ability to perform on unseen data by iteratively training and validating across different subsets of the dataset. Additionally, permutation feature importance was used to identify the most influential input features. This method evaluates the increase in prediction error when individual features are randomly permuted, offering insights into which synthesis parameters most significantly impact specific capacitance predictions.

The SVR model, illustrated in **Figure 2**a, achieves an RMSE of 0.14897, an MAE of 0.11359, and an R^2 of 0.41396, indicating a moderate fit between predicted and actual specific capacitance values. The corresponding scatter plot shows a close alignment of data points with the trend line, suggesting reasonable predictive accuracy. The RF model, shown in Figure 2b, yielded an RMSE of 0.13734, an MAE of 0.11054, and an R^2 of 0.35689. While RF demonstrates slightly lower error metrics compared to SVR, its marginally lower R^2 suggests that it captures slightly less variance in the data, despite effectively minimizing prediction errors. The neural network classifier (NNC), presented in Figure 2c, results in an RMSE of 0.15673, an MAE of 0.12398, and an R^2 of 0.35909. Although the error values are slightly higher, the scatter plot indicates a generally consistent alignment with actual values, reflecting moderate predictive capability. In contrast, the MLP model, depicted in Figure 2d, exhibited the poorest performance, with an RMSE of 1.0518, an MAE of 0.70824, and an R^2 of only 0.00975. The scatter plot reveals a wide dispersion of data points around the trend line, indicating a lack of predictive accuracy and poor model fit for specific capacitance prediction. Table S2, Supporting Information, presents the evaluation metrics for the ML models, highlighting that the SVR model achieves a moderate to best fit. It is characterized by relatively low error values and the highest R^2 score among the models, indicating its superior ability to explain variance in the data. The corresponding scatter plot in Figure 2 further supports this, showing a strong correlation between predicted and actual values. Figure S1, Supporting Information, provides a comparative overview of model performance, with error bars on RMSE illustrating the variability in training outcomes across models.

The NNC or ANN model exhibits moderately higher RMSE and MAE values, along with an R^2 score that is lower than SVR but comparable to RF. This suggests that while NNC captures certain trends in specific capacitance, its predictive accuracy is limited relative to SVR and RF. The observed pattern may reflect the

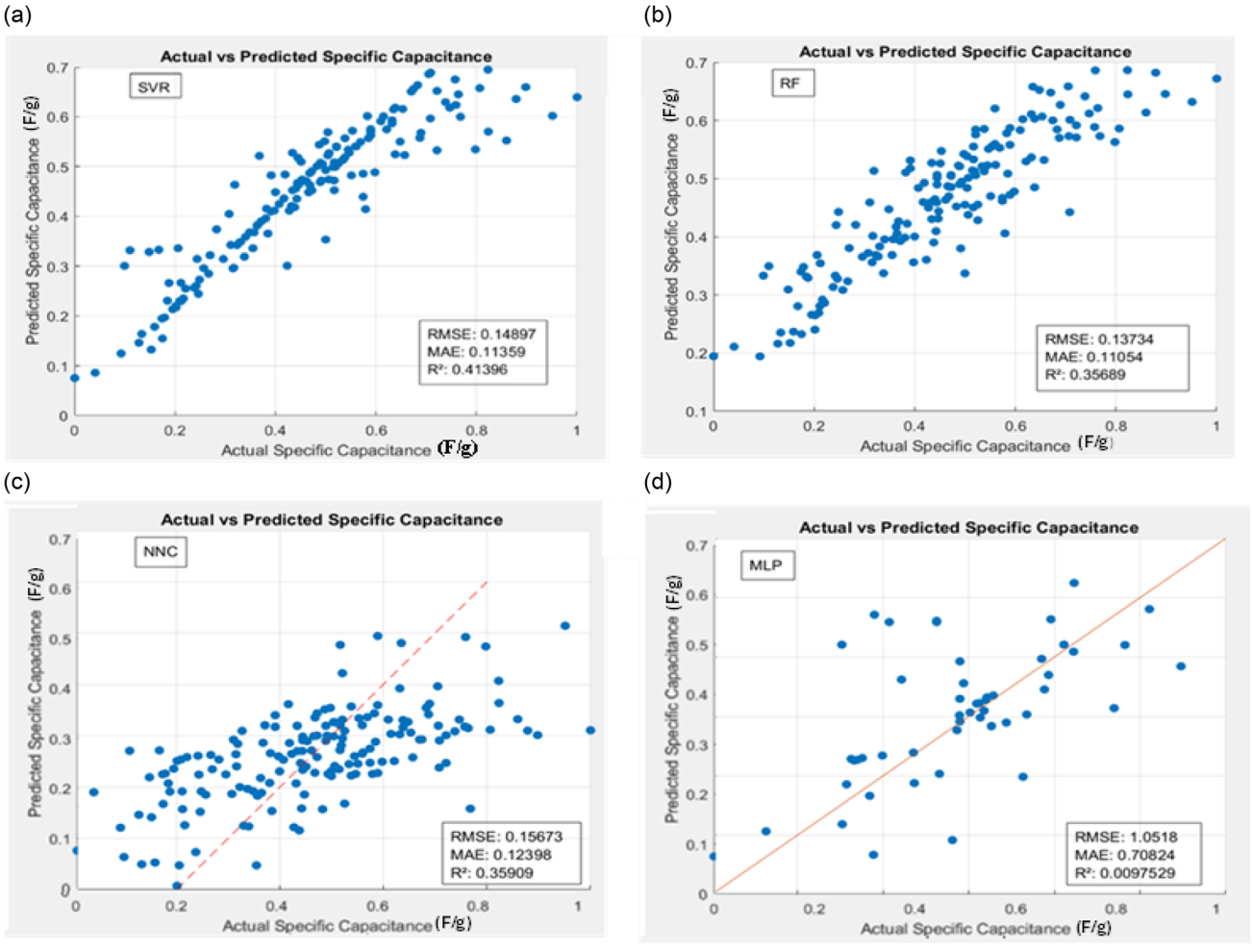

Figure 2. Comparison of the actual specific capacitance with predicted specific capacitance values of various models a) SVR; b) RF; c) NNC; and d) MLP.

model's insufficient adaptability to the complexity of the dataset, indicating potential underfitting or the need for further hyperparameter tuning. In contrast, the MLP model shows a significant deviation from the trend line, with the highest RMSE and MAE values and an extremely low R^2 score (as shown in Table S2, Supporting Information). This poor performance indicates that MLP fails to capture the underlying relationships between input features and specific capacitance. The widespread scatter of data points suggests limitations in the model's current configuration, possibly due to overfitting or inadequate training.

3.2. Analysis

Each ML model's predictive accuracy for specific capacitance was evaluated based on its alignment with the actual values and key performance metrics (RMSE, MAE, and R^2) as shown in Figure 2. Among the models, SVR demonstrated a well-balanced fit, characterized by relatively low RMSE and MAE values and a moderate R^2 score. This indicates that SVR effectively captures the underlying trend in specific capacitance and generalizes well to the dataset without significant overfitting or underfitting. The RF model also performed strongly, achieving the lowest RMSE and MAE values, which reflect high predictive accuracy. However, its slightly lower R^2 compared to SVR suggests that RF explains less of the overall variance, despite its effectiveness in minimizing prediction errors.

To further explore feature relationships, the Pearson correlation matrix in **Figure 3** reveals dependencies and potential multicollinearity within the dataset. A strong positive correlation between surface area and pore volume (0.8473) indicates that these features are closely linked, with larger pore volumes generally associated with increased surface area, both of which may collectively influence specific capacitance. Conversely, a notable negative correlation between activation temperature and activation time (–0.6725) suggests an inverse relationship, likely reflecting trade-offs in synthesis conditions. This inverse relationship reflects an optimization balance in the material preparation, where higher activation temperatures are often offset by lower carbonization temperatures and vice versa. Such trade-offs are common in fabrication processes aimed at achieving desirable porosity and surface functionality. Features such as activating agents, electrolytes, and voltage windows exhibit weaker correlations with other variables, suggesting that they contribute independently to specific capacitance without strong linear relationships. These weaker correlations enhance their predictive value, as they capture unique aspects of capacitance variation without overlapping effects.

The scatter plots in **Figure 4** illustrate the relationships between individual features and specific capacitance in supercapacitor performance. Figure 4a shows a weak, scattered correlation between biomass materials reported in the literature and specific capacitance, indicating minimal influence from intrinsic biomass properties. This suggests that most studies focus more on physicochemical and electrochemical characteristics than on the raw material itself.

Figure 4b reveals a high concentration of voltage windows data points within a narrow range, implying a limited impact on capacitance. Figure 4c demonstrates that surface area has an insignificant effect, as the points are widely spread without a discernible pattern, whereas Figure 4d indicates a more noticeable trend with current density, where higher current densities appear to stabilize capacitance values, suggesting a moderate correlation. Figure 4e shows a scattered relationship with pore volume, indicating no strong influence on capacitance.

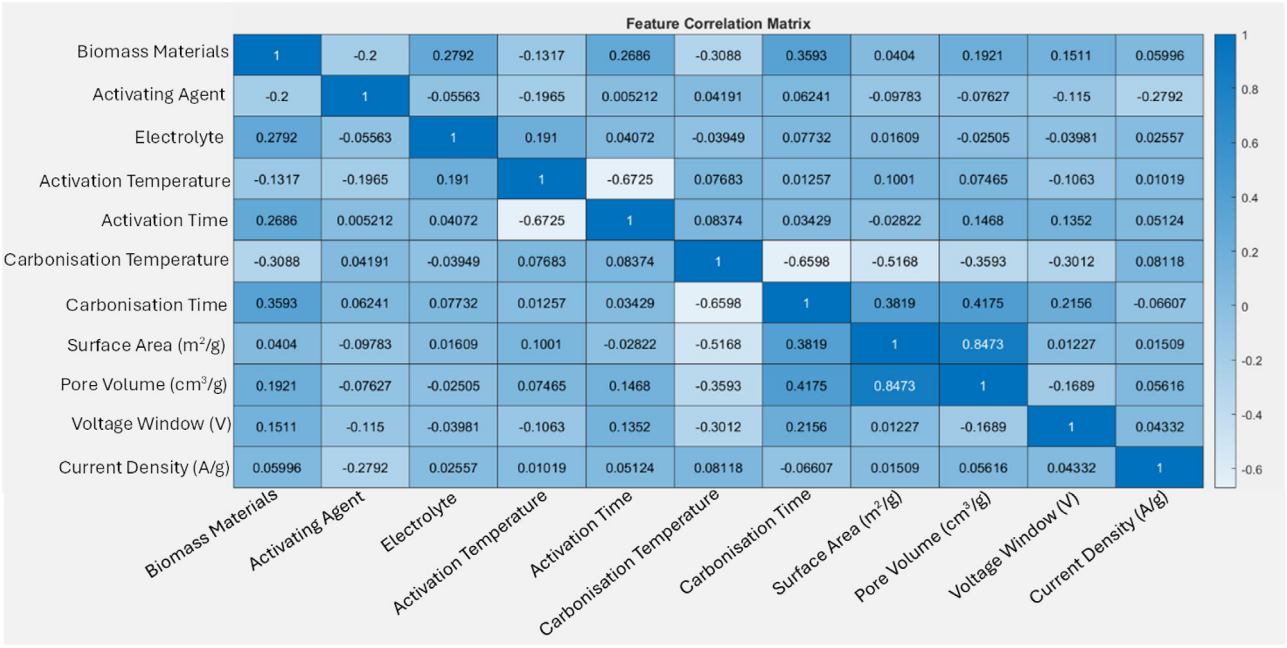

Figure 3. Feature correlation coefficient diagram by SVR model.

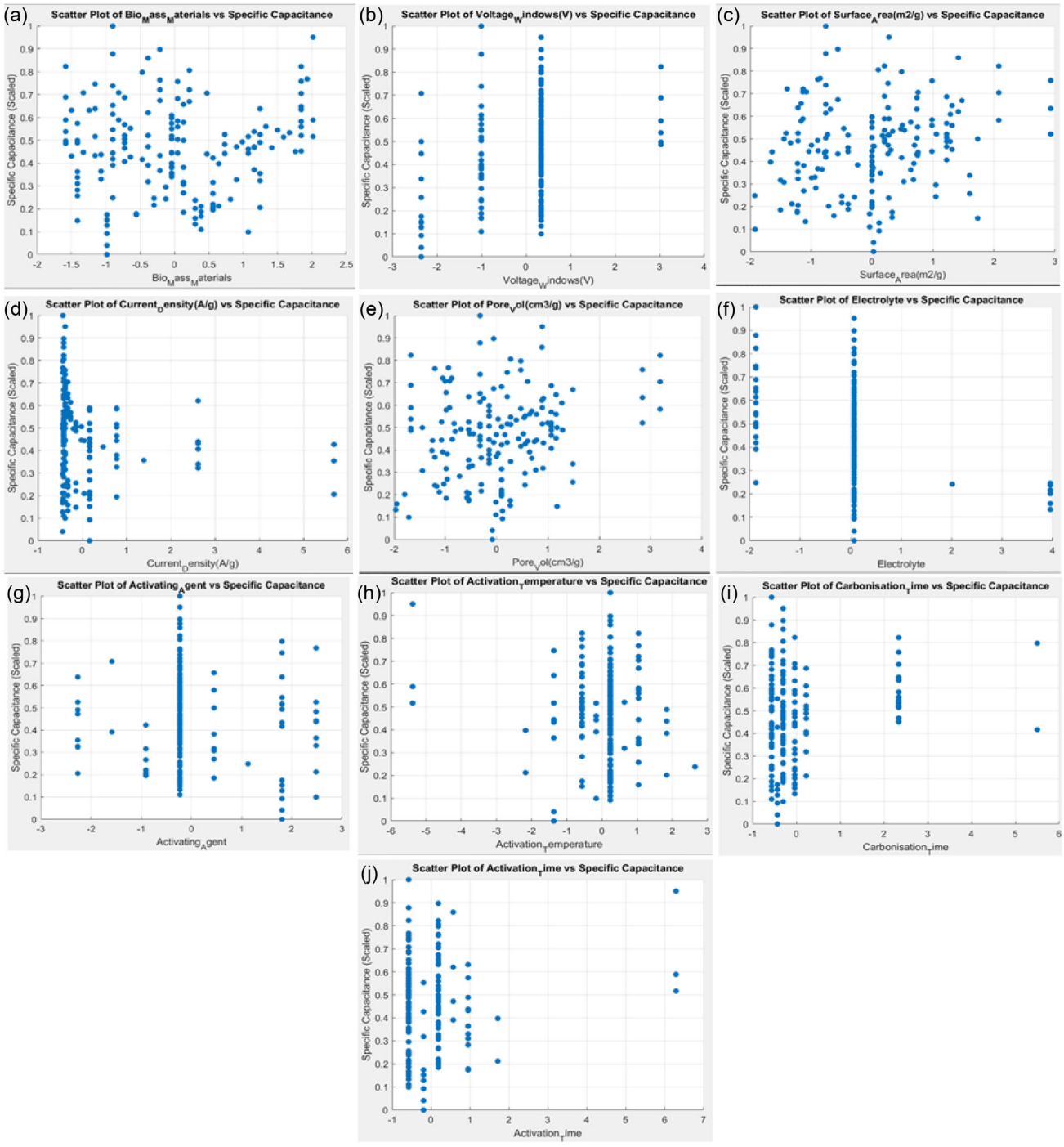

Figure 4. The correlation between the specific capacitance and materials properties, thermal treatment, and electrochemical factor of a) Biomass material; b) Voltage windows; c) Surface area; d) Current density; e) Pore volume; f) Electrolyte; g) Activating agent; h) Activation temperature; i) carbonization time; and j) activation time.

Figure 4f reveals that the majority of electrolyte data points cluster at a single level, implying minimal effect. Figure 4g displays that varying activating agent levels have little influence on capacitance, while Figure 4h shows that activation temperature remains largely uncorrelated, with data points clustering around zero. Figure 4i suggests that shorter carbonization time may allow for a broader range of capacitance values, but does not show a strong trend. Finally, Figure 4j indicates a slight influence of activation time, particularly at shorter durations, yet no clear linear trend is observed. Collectively, these observations suggest that specific capacitance is weakly correlated with most individual features, with current density showing the most consistent relationship. This highlights the importance of considering feature interactions and nonlinear effects in predictive modeling.

Figure 5 presents the permutation feature importance analysis for the two most effective ML models in this study, SVR and RF, highlighting the relative influence of each input feature on specific capacitance predictions. In the SVR model (Figure 5a), "biomass materials" emerged as the most influential feature, followed by "pore volume (cm^3/g)" and "voltage window (V)". These results suggest that the intrinsic properties of the biomass precursor, along with porosity and operating voltage range, play a significant role in determining specific capacitance. Features such as "carbonization temperature" and "current density ($A\,g^{-1}$)," showed moderate importance, indicating their partial contribution to performance outcomes. In contrast, activation time and carbonization time was among the least impactful features, suggesting limited influence on capacitance within the studied parameter range. Figure 5b presents the permutation feature importance rankings for the RF model, revealing notable differences compared to the SVR model. In RF, "Current Density (A/g)," "Surface Area ($m^2\,g^{-1}$)," and "Voltage Window (V)" emerge as the most influential features in predicting specific capacitance. While "Biomass Materials" also shows considerable importance, their influence is slightly reduced compared to their dominant role in the SVR model. Consistent with SVR, "activation time" and "carbonization time" rank among the least impactful features in the RF model. These results suggest that features such as "current density," "surface area," "pore volume," and "voltage window" are critical to capacitance prediction, although their relative importance varies across models. This variation reflects the underlying differences in model architecture: SVR is a kernel-based model that is sensitive to nonlinear relationships and can extrapolate trends even in sparse data regions. In contrast, RF is an ensemble of decision trees and excels at capturing interactions and thresholds. All four models were trained to predict specific capacitance, and their trends were compared to evaluate how each responds to changes in input features. This comparative analysis highlights the key factors influencing capacitance and provides valuable insights for optimizing supercapacitor design. According to model predictions, honeydew peel-derived carbon in a 1 M H_2SO_4 electrolyte, when doped with phosphorus-rich heteroatoms, significantly enhances specific capacitance, even at higher current densities. This improvement is attributed to increased electronic conductivity and enhanced electrode surface wettability, underscoring the importance of material selection and surface chemistry in supercapacitor performance.

3.3. Synthesis and Activation of Heteroatom-Doped Honeydew Peel

Several methods have been investigated for synthesizing AC from honeydew peel, as well as from peels of related fruits within the same botanical family, which have shown promise as electrode materials for supercapacitors.[17,28–32] In this study, we employed a straightforward synthesis approach. The powdered honeydew peel was first subjected to carbonization in a muffle furnace at 300 °C, with a heating rate of 2.5 °C min^{-1} for 4 h. The resulting carbonized material was then thoroughly mixed with an optimized concentration of phosphoric acid (H_3PO_4) as the activating agent. The resulting slurry was oven-dried and subsequently activated in a muffle furnace at 500 °C, maintaining the same heating rate for 1 h. After activation, the sample was allowed to cool to room temperature. Phosphoric acid was selected due to its ability to generate mesopores, thereby enhancing both pore diameter and volume. Additionally, its use simplifies the recovery of the carbon product, requiring only a water rinse. This efficiency is attributed to the relatively low reaction temperatures and the cross-linking interactions between H_3PO_4 and the honeydew

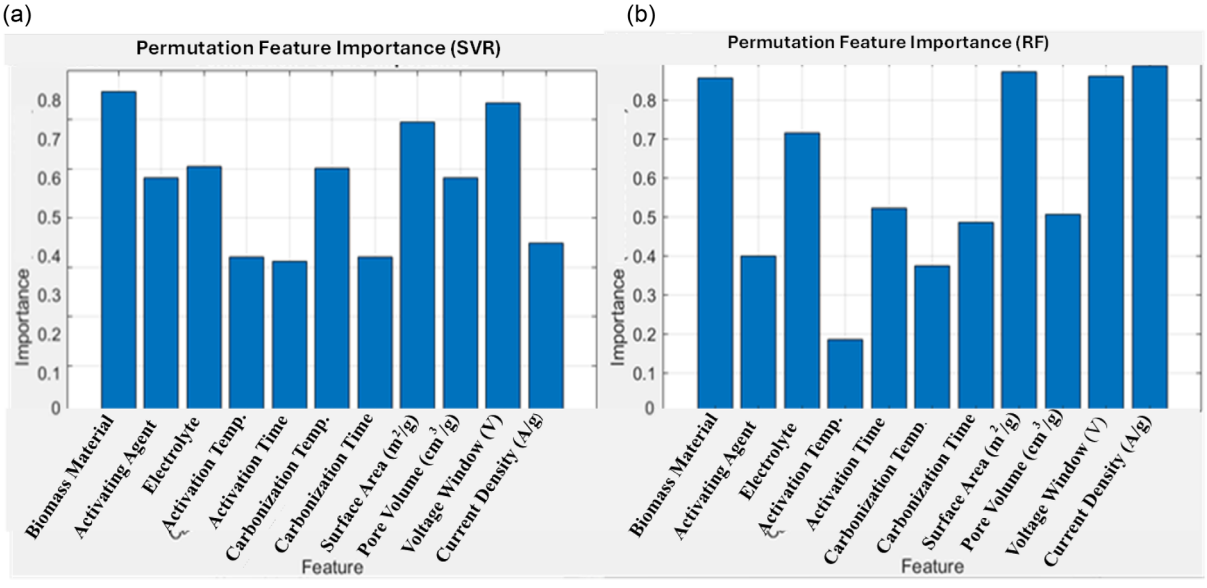

Figure 5. Permutation feature importance analysis of SVR and RF.

peel, which inhibit the release of volatiles and the formation of tar during pyrolysis.[33] The resulting heteroatom-doped honeydew peel carbon, synthesized at an optimized precursor-to-activating agent ratio of 1:4, presents a promising pathway toward achieving the theoretical capacitance limits of porous carbon materials.

3.4. Physicochemical Properties of Heteroatom-Doped (oxygen and Phosphorous-Rich) Honeydew Peel Material

The honeydew peel sample was finely ground into a powder to ensure a smooth and homogenous surface for X-ray diffraction analysis. The X-ray diffraction pattern of the AC derived from honeydew peel (**Figure 6**a) exhibited characteristic peaks around 25° and 44° corresponding to the (002) plane of turbostratic carbon (t-carbon)[29] and the (101) plane of graphitized carbon,[17] respectively. Complementary Raman spectroscopy (Figure 6b) revealed distinct peaks at approximately 1350 cm^{-1} (D-band, indicating structural disorder), 1600 cm^{-1} (G-band, representing graphitic domains), and around 2700 cm^{-1}, confirming the formation of AC. The presence of these peaks reflects the intrinsic molecular structure of amorphous carbon in the honeydew peel. The intensity ratio of the D to G bands (I_D/I_G) was calculated to be 1.05, suggesting a high degree of defects in the graphitic structure, an advantageous feature for energy storage applications. The activation process begins with cellulose depolymerization, followed by dehydration of biopolymers to form aromatic rings, ultimately leading to the development of porous carbon.[33,34] During this process, water-soluble KCl, formed during carbonization, may react with the honeydew peel under thermal treatment. Chlorine is released as a gas, while potassium melts and integrates into the carbon matrix, contributing to the formation of porous graphitic carbon. This structural transformation is further supported by SEM images (**Figure 7**a,b), which reveal a compact and porous architecture conducive to efficient ion transport.

The use of H_3PO_4 as an activating agent plays a critical role in pore development during the early stages of activation.[34,35] The macroporous open structure observed in Figure 7b, resulting

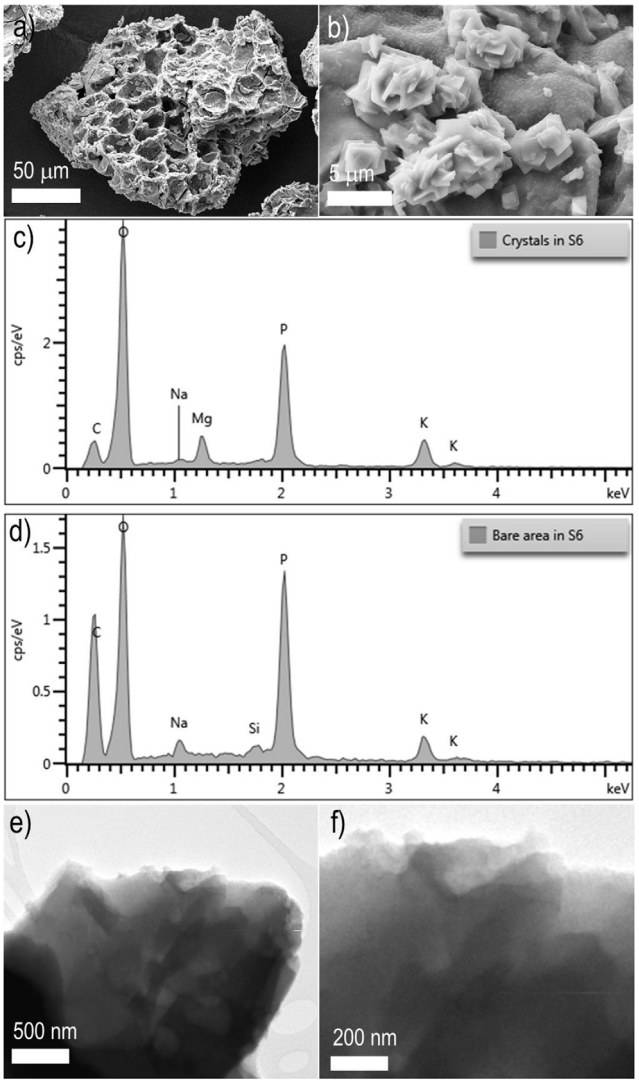

Figure 7. a,b) FE-SEM images, c,d) and the corresponding EDS spectra confirming the presence of oxygen and phosphorus species, e,f) TEM images of the honeydew peel-derived porous carbon.

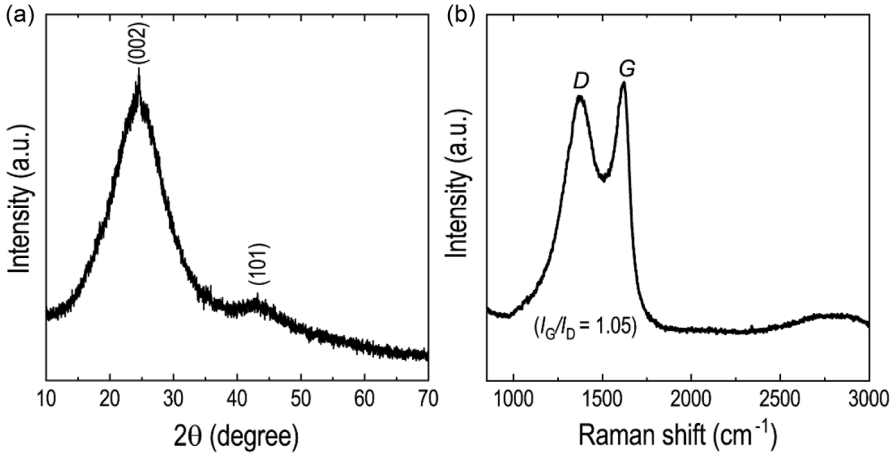

Figure 6. a) X-ray diffraction pattern and b) Raman spectra of the HDP-AC. G denotes graphitic and D denotes defect bands in the spectra.

from intense cross-linking, is well-suited for high-rate capability and serves as a buffer zone for rapid ion diffusion. Additionally, the cascaded crystal particles visible in Figure 7b are likely phosphorus deposits on the surface of the AC. Energy-dispersive X-ray spectroscopy (EDS) data (Figure 7c,d) confirm the presence of oxygen- and phosphorus-rich species, with the O/P atomic ratio indicating co-doping of these elements into the carbon matrix. TEM images (Figure 7e,f) further validate the porous nature and amorphous regions of the synthesized carbon material.

3.5. Electrochemical Properties of Heteroatom-Doped Honeydew Peel Material

The electrochemical energy storage properties of heteroatom-doped honeydew peel-derived porous carbon were initially assessed using CV at scan rates ranging from 5 to 50 mV/s, and galvanostatic charge–discharge (CD) measurements at various current densities. These tests were conducted in a conventional three-electrode system with 1 M H_2SO_4 as the electrolyte. The corresponding results are presented in **Figure 8**a,b. Both the CV and CD curves exhibited nearly rectangular and triangular shapes, respectively, characteristic of ideal EDLC behavior, indicating excellent capacitive performance and high specific capacitance reaching up to 485 F g^{-1}.

The enhanced specific capacitance is attributed to the presence of oxygen- and phosphorus-rich species, which facilitate the formation of P—O—C structures through cross-linking and oxidation reactions. This structural evolution, combined with the preservation of micropores as observed in microscopy images, contributes to superior charge storage capability. Encouraged by the excellent performance within a 1 V voltage window, a symmetric supercapacitor device was fabricated using a two-electrode configuration with 1 M H_2SO_4 as the electrolyte. The CV and CD profiles of the symmetric device (**Figure 9**a,b) closely resemble those of the three-electrode setup, confirming EDLC behavior driven by ion adsorption on the porous carbon electrodes.

Notably, the area under the CV curve in Figure 9a is significantly larger than that of the single electrode in Figure 8a, reflecting a higher specific capacitance. The symmetric device achieved a specific capacitance of 611 F g^{-1} at a current density of 1.3 A g^{-1} (Figure 9b) and delivered a maximum energy density of 35 Wh kg^{-1} at a power density of 650 W kg^{-1}. This performance surpasses that of most reported carbon-based symmetric capacitors (see Table S1, Supporting Information), which typically rely on materials with high surface area and pore volume. In contrast, our material, phosphorus and oxygen co-doped honeydew peel-derived carbon, exhibits a relatively low surface area of 33 $m^2 g^{-1}$. Across our study, the samples demonstrated a wide range of Brunauer-Emmett-Teller surface areas, from 224 down to 3.0049 $m^2 g^{-1}$, with corresponding pore volumes ranging from 0.076 to 0.000867 $cm^3 g^{-1}$. These variations are attributed to the differing concentrations of the activating agent used, from low to high levels. However, under these conditions, both the SVR and RF models predicted relatively low specific capacitance values, 229 F g^{-1} and 208 F g^{-1}, respectively. This discrepancy may be explained by several contributing factors[36,37] 1) insufficient surface area and pore volume for effective ion adsorption. 2) A limited voltage window restricting energy storage. 3) moderate current density affecting ion transport.

These predictions suggest that under suboptimal physical conditions, such as low surface area and pore volume, the ML models estimate low to moderate specific capacitance values. This is expected, as these features are strongly correlated with ion accessibility and charge storage capacity. However, in real experimental settings, additional surface properties often play a significant role in enhancing charge storage, leading to much higher capacitance values, such as the observed 611 F g^{-1}. The discrepancy between model predictions and experimental results may be attributed to phosphorus and oxygen co-doping, which can significantly enhance electrochemical performance by introducing pseudocapacitive behavior and improving conductivity. This highlights the critical importance of validating model predictions with actual experimental data, especially when material-level modifications are involved.

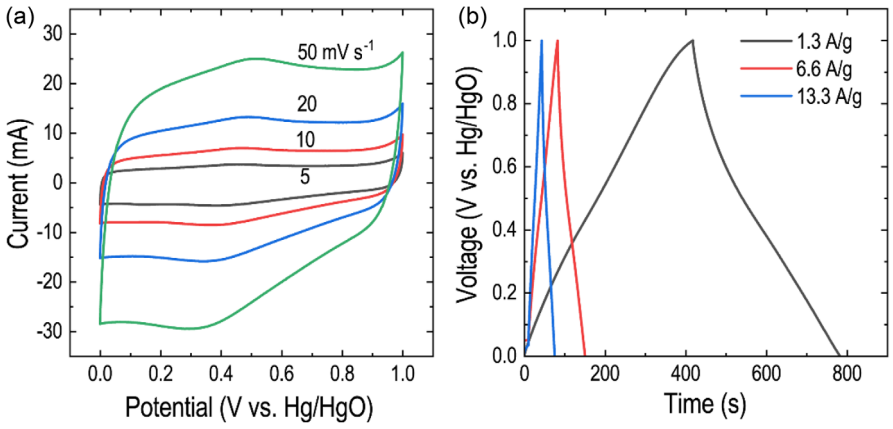

Figure 8. a) CV, and b) charge–discharge (CD) curves for the AC sample derived from the honeydew peel tested in a three-electrode configuration in 1M H_2SO_4 electrolyte.

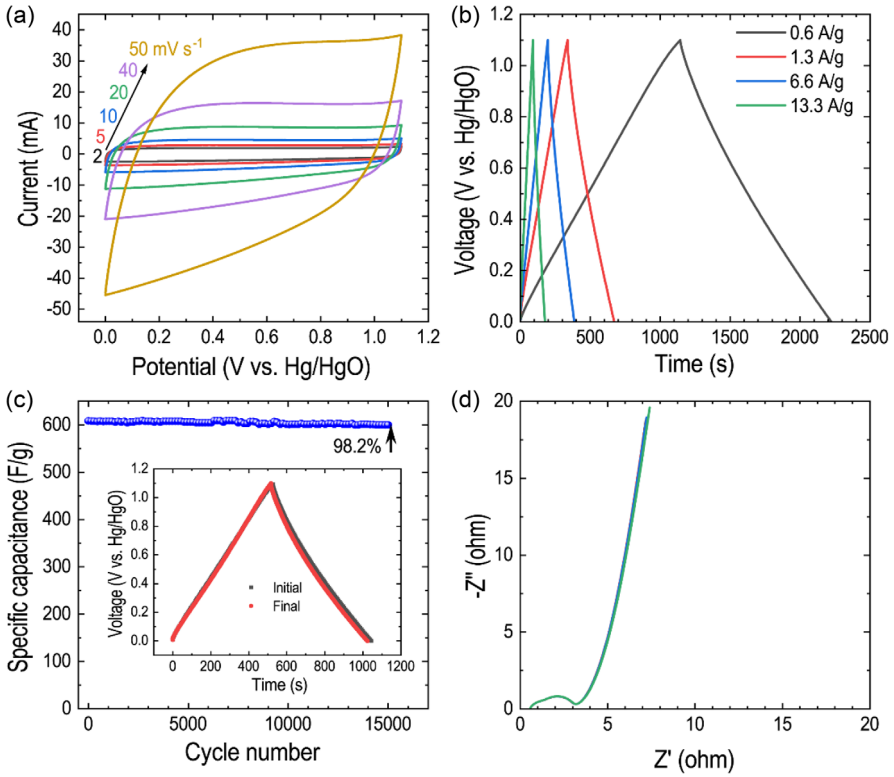

Figure 9. a) CV, b) charge–discharge (CD) curves for the AC sample derived from the honeydew peel tested in a two-electrode configuration as a symmetric supercapacitor device in 1MH_2SO_4 electrolyte. c) cycling stability of the device with the inset showing the charge–discharge curves superimposed for initial and final cycles, and the corresponding d) Nyquist plots showing no visible changes in the shape of the curve.

The long-term cycling performance of the biomass-derived symmetric supercapacitor, shown in Figure 9c, demonstrates excellent stability, retaining 98.2% of its initial specific capacitance after 15,000 charge–discharge cycles at a current density of 1.3 A g^{-1}. EIS results, presented in Figure 9d, reveal that the Nyquist plot remains nearly unchanged after extended cycling, indicating consistent resistance behavior and robust electrode integrity.

Overall, the Nyquist plot exhibits features characteristic of an ideal capacitor: a nearly vertical line at high frequencies, representing capacitive behavior, and a 45° slope at mid frequencies, attributed to the distributed resistance within the porous carbon electrode. These experimental findings provide valuable insights into the electrochemical behavior of the material and support the predictive capabilities of ML models. The input variables that most significantly influence performance include the nature of the biomass precursor, the optimized concentration of the activating agent, and the choice of electrolyte.

3.6. Discussions and Limitations of the Current Study

This study highlights the importance of narrowing the scope of data collection to minimize variability in the analysis of supercapacitor performance. Focusing on specific types of biomass materials, particularly those that are commonly utilized or possess similar properties (e.g., cellulose content), is essential to reduce variations associated with diverse starting materials.[28] Additionally, outliers, especially those reporting unusually high specific capacitance values, should be critically evaluated and potentially excluded if they demonstrate inconsistencies with other relevant parameters. When available, pore size distribution data should be incorporated with caution, ensuring consistency in measurement techniques, as pore architecture plays a critical role in determining capacitance.[29] Standardizing these variables enhances the reliability of comparative analyses and improves the predictive accuracy of ML models applied to supercapacitor research.

It is essential to distinguish between carbonization and hydrothermal carbonization (HTC) when collecting data for ML applications in supercapacitor research. These two processes yield carbon materials with distinct physical and chemical characteristics that significantly influence key performance metrics such as specific capacitance, surface area, and pore structure. For instance, honeydew peel-derived carbon produced via conventional carbonization exhibits notable differences in surface area, pore size distribution, and surface functional groups compared to materials synthesized through HTC.[17,34]

These discrepancies directly affect the suitability and performance of the resulting materials in energy storage applications. Moreover, the operational conditions, such as temperature, pressure, and reaction medium (dry vs. wet), differ substantially

between carbonization and HTC, leading to unique reaction pathways and end products. These variations influence carbon yield, structural integrity, and purity, all of which contribute to electrochemical performance differences.

Consequently, materials produced via HTC (hydrochar) may exhibit electrochemical behaviors that diverge from those of conventionally carbonized materials (biochar). For ML applications, it is therefore imperative to treat carbonization and HTC as distinct processes. By incorporating the synthesis method as a separate feature in the ML model, the algorithm can more accurately capture the influence of each process on performance outcomes, such as specific capacitance.[30]

The choice of activating agents, such as KOH, $CaCl_2$, K_2FeO_4, $MgCl_2$, H_3PO_4, and steam, plays a pivotal role in determining the physicochemical properties of the resulting carbon materials. While all these agents serve to activate carbon, they operate through distinct mechanisms, leading to significant variations in surface area, pore architecture, and chemical composition. Therefore, treating these agents as equivalent during data collection for ML applications is inappropriate and may compromise model accuracy. Chemical activation, involving agents like KOH, H_3PO_4, $CaCl_2$, and K_2FeO_4, induces specific reactions with the carbon precursor, influencing pore formation, surface area, and the incorporation of functional groups.[31,34] Each agent produces unique pore size distributions and chemical functionalities. In contrast, physical activation, typically using steam or $MgCl_2$, relies on gasification and controlled carbon burn-off at elevated temperatures to enhance porosity. Some agents, such as K_2FeO_4, may also exhibit catalytic effects, further modifying the carbon structure and its electrochemical behavior.[32]

The selected activating agent significantly affects performance metrics such as specific capacitance, energy density, and cycling stability. For example, KOH-ACs are often favored for their high surface area and conductivity, while H_3PO_4-ACs are valued for their stability in aqueous electrolytes. For ML applications, it is essential to encode each activating agent as a distinct categorical feature. This enables the model to learn the individual and interactive effects of each agent on material properties and performance. Additionally, recording process parameters, such as activation temperature, duration, and carbonization conditions, alongside the activating agent, ensures consistency and enhances model reliability.

It is also important to acknowledge that the data used in this study were sourced from various research articles (listed in Table S1), each conducted under different environmental and experimental conditions. Such diversity may introduce inconsistencies in ML outcomes. Furthermore, variations in terminology across studies can lead to misinterpretations and compromise the integrity of the final analysis. The failure to account for differences in carbonization methods, biomass types,[37] and activation techniques may have influenced the appropriateness of the study's results. Therefore, future researchers are advised to exercise caution and rigor in data collection, ensuring that these critical factors are carefully considered and consistently documented.

4. Conclusions

This study presents an ML-guided investigation into oxygen and phosphorus-rich heteroatom-doped carbon material derived from biomass for supercapacitor applications. The aqueous symmetric supercapacitor demonstrated a strong correlation between specific capacitance and the nature of biomass precursor, a finding that stands out as particularly novel. By integrating a diverse dataset compiled from various published sources, the research effectively highlights the complex interrelationships among critical features such as biomass type, surface area, pore volume, and experimental parameters, including activation and carbonization temperatures.

The performance evaluation of four distinct ML models, SVR, RF, ANN, and MLP, demonstrated their predictive capabilities for supercapacitor performance. Among these, the SVR model emerged as the most effective, achieving a coefficient of determination (R^2) of 0.41398 and a MAE of 0.11359. These findings underscore the potential of ML to enhance the understanding and optimization of biomass-derived carbon materials for energy storage.

Importantly, the study identified surface area and pore volume as key contributors to specific capacitance. However, in the case of heteroatom-doped honeydew peel-derived carbon, the influence of biomass composition and dopant chemistry appeared to outweigh the effects of surface area and porosity. This insight provides a valuable foundation for future research aimed at tailoring carbon material properties for enhanced energy storage performance. The study also emphasizes the importance of focused data collection and rigorous consideration of material preparation methods. These recommendations are critical for improving the reliability of ML models and advancing the field of supercapacitor research. Experimentally, the symmetric supercapacitor fabricated in this work achieved a specific capacitance of 611 $F\,g^{-1}$, with an energy density of 35 $Wh\,kg^{-1}$ and a power density of 650 $W\,kg^{-1}$, performance metrics that surpass many reported carbon-based systems. Overall, this work contributes meaningfully to the growing body of knowledge in energy storage, advocating for the continued integration of ML techniques to accelerate the development of high-performance, sustainable supercapacitor materials.

Acknowledgements

M.M. acknowledges the Winston Churchill Fellowship to learn globally and inspire locally. This work was partially supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (grant Nos. JP20H00392 and JP23H05459).

Open access publishing facilitated by Murdoch University, as part of the Wiley - Murdoch University agreement via the Council of Australian University Librarians.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: biomass · carbon · dopant · energy · machine learning · storage

[1] G. G. Prasad, N. Shetty, S. Thakur, Rakshitha, K. B. Bommegowda, in *IOP Conf. Series: Materials Science and Engineering*, IOP Publishing Ltd., *561*, **2019**, p. 012105.
[2] S. Butler, *Supercapacitors vs. Batteries: What's the Difference?* https://www.howtogeek.com/786195/supercapacitors-vs-batteries-whats-the-difference/#:~:text=Supercapacitors%20have%20been%20around%20since%20the%201950s.**2022**, (accessed: July 2025).
[3] M. A. Dar, S. R. Majid, M. Satgunam, C. Siva, S. Ansari, P. Arularasan, S. R. Ahamed, *Int. J. Hydrogen Energy* **2024**, *70*, 10.
[4] L. Wang, X. Hu, *Chem. Asian J.* **2018**, *13*, 1518,
[5] S. Rawat, R. K. Mishra, T. Bhaskar, *Chemosphere* **2022**, *286*, 131961.
[6] Y. Liu, J. Chen, B. Cui, P. Yin, C. Zhang, *C J. Carbon Res.* **2018**, *4*, 53.
[7] X. Yuebin, X. Liu, W. Xiong, H. Wang, X. Ji, F. Kong, G. Yang, J. Xu, *Ind. Crops Prod.* **2021**, *174*, 114184.
[8] J. Li, R. Xiao, M. Li, H. Zhang, S. Wu, C. Xia, *Fuel Proces. Technol.* **2019**, *192*, 239.
[9] D. Zhang, P. Yang, Y. Zhang, H. Liu, Y. Xu, J. Wu, P. Li, *J. Energy Storage* **2022**, *56*, 106134.
[10] V. Sawant, R. Deshmukh, C. Awati, *J. Energy Chem.* **2023**, *77*, 438.
[11] J. Jia, Z. Yao, L. Zhao, T. Xie, Y. Sun, L. Tian, L. Huo, Z. Liu, *Biomass Bioenergy* **2023**, *175*, 106858.
[12] S. Koutcheiko, V. Vorontsov, *J. Biobased Mater. Bioenergy* **2013**, *7*, 733.
[13] Z. Dong, X. Bai, D. Xu, W. Li, *Bioresour. Technol.* **2023**, *367*, 128182.
[14] L. Leng, L. Yang, X. Lei, W. Zhang, Z. Ai, Z. Yang, H. Zhan, J. Yang, X. Yuan, H. Peng, H. Li, *Biochar* **2022**, *4*, 63.
[15] X. Yang, C. Yuan, S. He, D. Jiang, B. Cao, S. Wang, *Fuel* **2023**, *331*, 125718.
[16] Y. Sun, P. Sun, J. Jia, Z. Liu, L. Zhao, Y. Zhao, W. Niu, Z. Yao, *Chem. Eng. J.* **2024**, *485*, 149975.
[17] M. Minakshi, A. Samayamanthry, J. Whale, R. Aughterson, P. A. Shinde, K. Ariga, L. K. Shrestha, *Chem. Asian J.* **2024**, *19*, e202400622.
[18] A. Seraj, M. M. Khanaposhtan, R. Daneshfar, M. Naseri, M. Esmaeili, A. Baghban, S. Habibzadeh, S. Eslamian, *Handbook of Hydroinformatics*, Elsevier **2023**, pp. 89–105.
[19] A. Jamaluddin, D. Harjunowibowo, S. Budiawanti, N. A. Kurdhi, S. D. T. Lai, S. Ramesh, *Energy Rep.* **2023**, *10*, 3125.
[20] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova, C. Zhong, *Statistic Surveys* **2022**, *16*, 1.
[21] I. Tsochantaridis, T. Hofmann, T. Joachims, Y. Altun, in Proceedings of the twenty-first int. conf. on Machine learning **2004**, p. 104.
[22] I. Turk, *Practical MATLAB*, Apress, Berkeley, CA **2019**.
[23] Z. M. Omer, H. Shareef, *Intell. Syst. Appl.* **2023**, *19*, 200261.
[24] R. S. Abirami, G. S. Kumar, *SN Comput. Sci.* **2022**, *3*, 1.
[25] A. E. Eldeen, M. A. Azim, M. Abdelsattar, A. AbdelMoety, *J. Energy Storage* **2024**, *100*, 113556.
[26] S. A. N. Alexandropoulos, A. S. B. Kotsiantis, M. N. Vrahatis, *Knowl. Eng. Rev.* **2019**, *34*, e1.
[27] N. Redell, (Preprint) arXiv: 1908.09718 **2019**, 09718, https://arxiv.org/abs/1908.09718.
[28] Z. Wang, P. Tammela, M. Strømme, L. Nyholm, *Adv. Energy Mater.* **2017**, *7*, 1700130.
[29] T. Wang, R. Pan, M. L. Martins, J. Cui, Z. Huang, B. P. Thapeliya, C.-L. Do-Thanh, M. Zhou, J. Fan, Z. Yang, M. Chi, T. Kobayashi, J. Wu, E. Mamontov, S. Dai, *Nat. Commun.* **2023**, *14*, 4607.
[30] M. Aghaaminiha, R. Mehrani, S. Sharma, *Biomass Convers. Biorefin.* **2023**, *13*, 9855.
[31] X. Pang, M. Cao, J. Qin, X. Li, X. Yang, *J. Porous Mater.* **2022**, *29*, 559.
[32] D. M. Demir, M. Doguscu, *ChemistrySelect* **2022**, *7*, e202104295.
[33] M. A. Yahya, Z. Al-Qodah, C. W. Z. Ngah, *Renewable Sustainable Energy Rev.* **2015**, *46*, 218.
[34] M. Fan, Y. Shao, Y. Wang, J. Sun, H. He, Y. Guo, S. Zhang, S. Wang, B. Li, X. Hu, *Renew. Energy* **2024**, *240*, 122151.
[35] T. Adinaveen, L. J. Kennedy, J. J. Vijaya, G. Sekaran, *J. Mater, Cycles Waste Manag.* **2015**, *17*, 736.
[36] L. Zhu, H. Yin, L. Ju, B. Zhou, B. Hu, F. Hou, Y. Xie, J.-L. Zhan, W. Du, *Ceram. Int.* **2024**, *50*, 48609.
[37] W. Du, Z. Zhang, L. Du, X. Fan, Z. Shen, X. Ren, Y. Zhao, C. Wei, S. Wei, *J. Alloys Compd.* **2019**, *797*, 1031.

Manuscript received: May 20, 2025
Revised manuscript received: July 13, 2025
Version of record online: August 29, 2025