

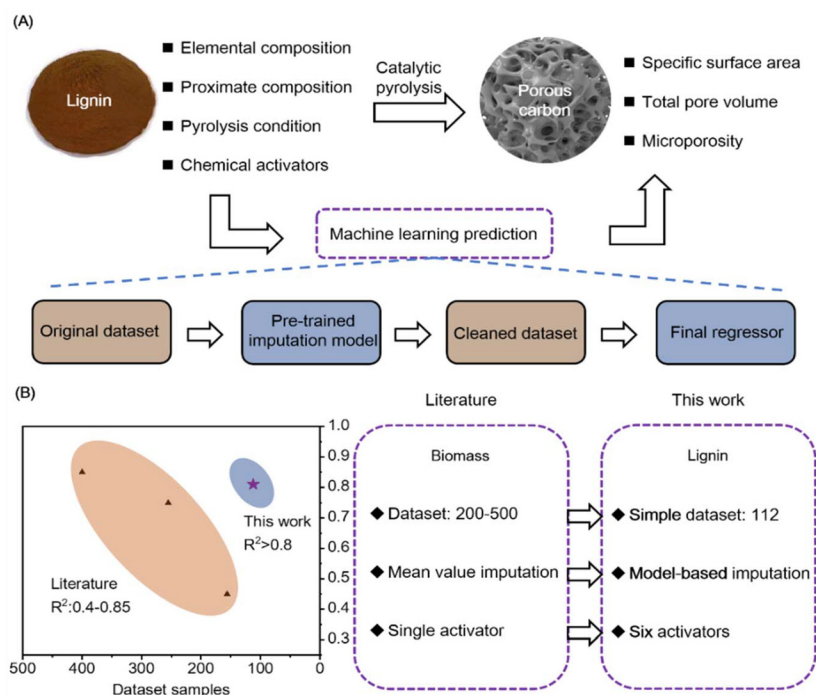


Cite this: *Green Chem.*, 2025, **27**, 2046

Machine learning prediction of physical properties of lignin derived porous carbon *via* catalytic pyrolysis†

Zihao Xie, Yue Cao* and Zhicheng Luo  *

Lignin-derived porous carbon produced through catalytic pyrolysis is crucial for energy storage, adsorption, and catalysis. However, predicting specific surface area (SSA), total pore volume (TPV), and microporosity (MP) remains challenging due to the variability in lignin properties, chemical activators, and pyrolysis conditions, compounded by limited data availability. In this study, we applied a hybrid machine learning framework incorporating a pre-trained interpolation model and a final regressor to impute missing features, improving prediction accuracy and generalizability. This approach yielded high predictive accuracy with R^2 values of 0.82



important structural features that are related to the elemental composition and properties of LDPC. However, due to the limited structural data available in the literature, these features were not included in this study. The chemical activators encompassed the type of activating agent (Agent) and the impregnation ratio (A/S). The pyrolysis conditions included pyrolysis temperature (Temp), retention time (RT), and heating rate (HR).

To ensure consistency in the dataset units, VM, Ash, FC, C, H, O, N, and S were all measured in %. SSA was measured in $\text{m}^2 \text{g}^{-1}$, TPV in $\text{cm}^3 \text{g}^{-1}$, Temp in $^\circ\text{C}$, RT in hours, and HR in $^\circ\text{C min}^{-1}$. Agent includes the following six type: KOH, ZnCl_2 , H_3PO_4 , K_2CO_3 , NaOH, Na_2CO_3 . In total, we collected 112 samples (*i.e.*, 112 porous carbon) from 32 published papers to predict the specific surface area (SSA), total pore volume (TPV), and microporosity (MP) of LDPC prepared *via* catalytic pyrolysis (Table S2†).

Dataset normalization and analysis

Model interpretation

Despite their ability to capture complex, nonlinear relationships and achieve high predictive accuracy, interpreting machine learning models remains challenging. We employed several techniques to enhance the interpretability of our final regressor.

Permutation importance analysis. This method aids in understanding the contribution of each input feature to the model's predictions. It involves systematically shuffling the values of individual features, retraining the model, and then measuring the change in predictive performance. A higher permutation importance score indicates that the feature has a more significant impact on the target variable, while scores close to zero suggest minimal influence.

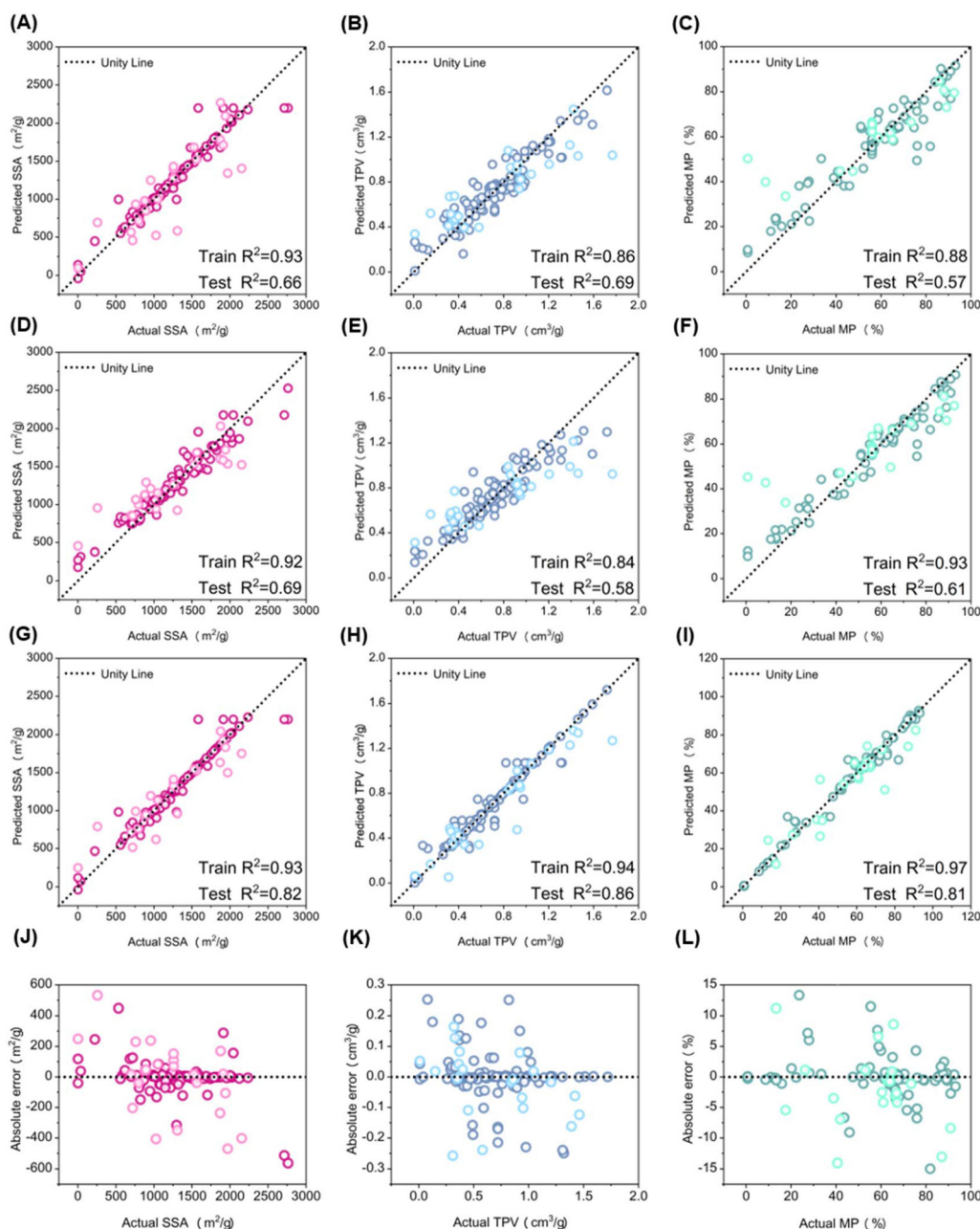
Partial dependence plots. To further elucidate the relationship between input features and the target variable, we utilized Partial dependence plots. These plots illustrate how changes in a single feature affect the predicted outcome, while keeping all other features constant. By visualizing these dependencies, PDPs provide insights into the direction and magnitude of a feature's influence on the model's predictions. This approach enhances interpretability by revealing trends and interactions that might not be immediately evident from raw data analysis or permutation importance alone.

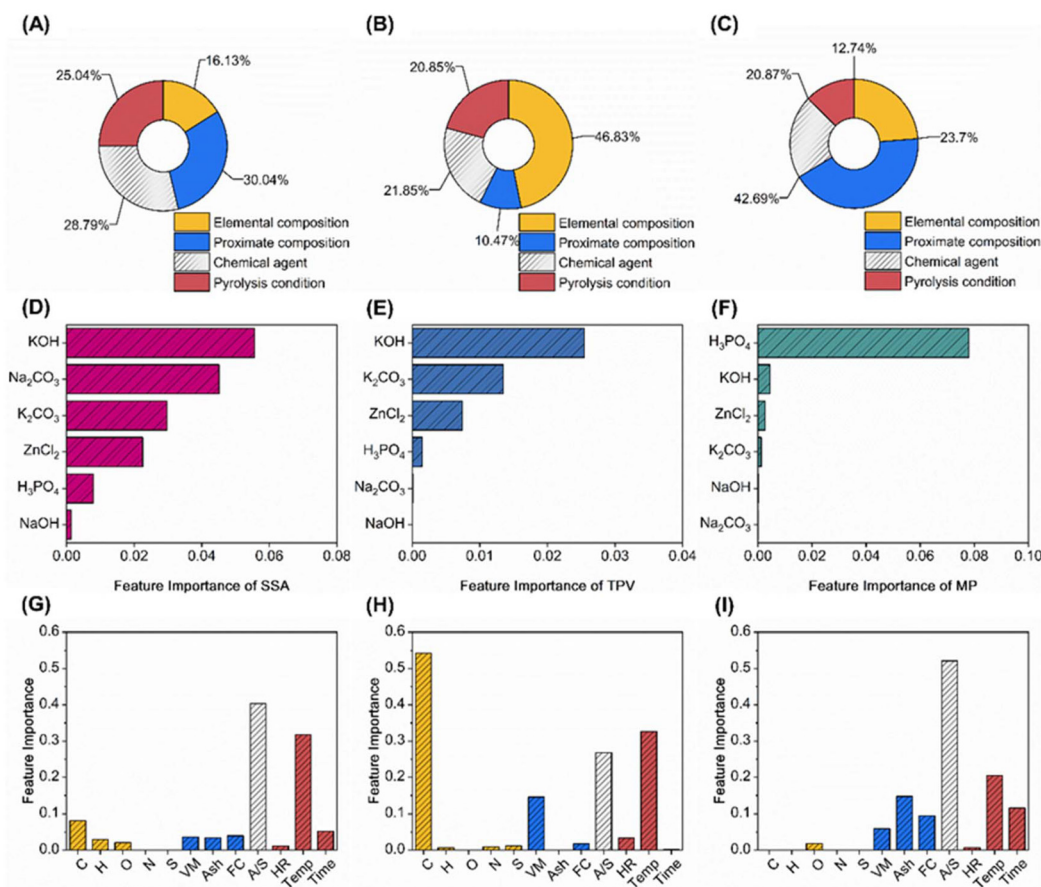
By integrating permutation importance analysis and partial dependence plots, we gained a comprehensive understanding of how specific features, such as chemical activators and pyrolysis conditions, influence the SSA, TPV, and MP of LDPC synthesized *via* catalytic pyrolysis. These interpret



Fig. 2 Pearson correlation matrix among any two features of the cleaned dataset.

GBR). To evaluate its performance, we also trained traditional Gradient Boosting Regressor (GBR) and Random Forest (RF) models on the original dataset to predict the same targets: SSA





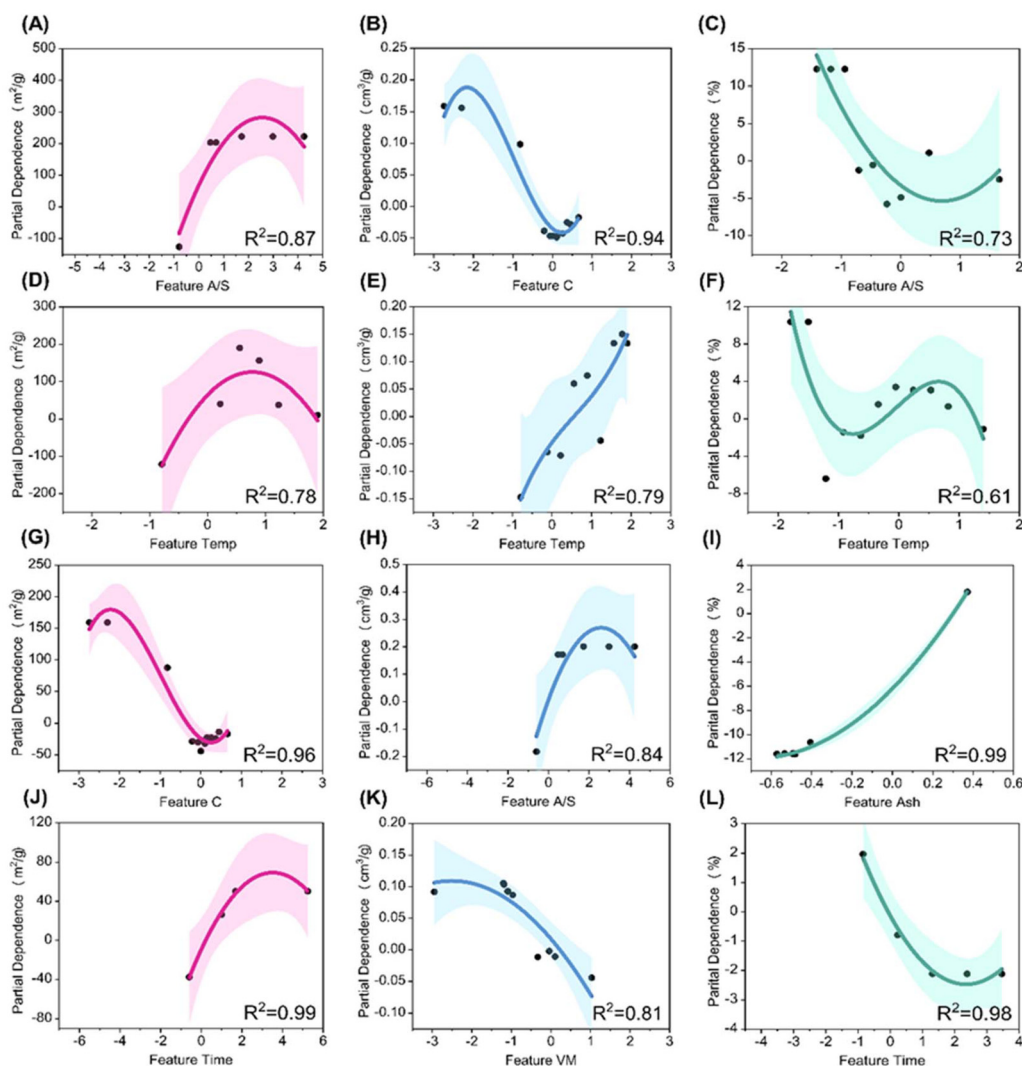


Fig. 5 Partial dependence analysis of top four input features on each target: SSA of LDPC with KOH (A, D, G and J); TPV of LDPC with KOH (B, E, H and K); and MP of LDPC with H_3PO_4 (C, F, I and L).

carbons synthesized under KOH activation (Fig. 5E), with TPV tending to increase as the temperature rises, indicating a transition from micropores to mesopores with higher temperatures. The trend of Temp and Time affecting SSA and TPV under KOH is similar to findings by Li *et al.* found that higher pyrolysis temperatures increase both SSA and TPV, and increasing pyrolysis time from 1 h to 2 h led to higher SSA, consistent with our machine learning model.³⁷

Regarding the influence of differences in elemental and proximate composition on SSA, TPV, and MP, Fig. 5G and K highlights VM and C as the most significant features. VM is decomposed to small gas molecules to physically create pores. This physical stripping is often uncontrollable, which has a lower pore formation efficiency compared to that of catalytic pyrolysis. Therefore, with an increase in VM, TPV tends to decrease (Fig. 5K). The negative correlation between VM and TPV is also supported by previous studies.^{37,38} For the element C, as the activation agent mainly etches carbon atoms to form pore structures, fewer carbon atoms are more easily etched into pores. Conversely, a higher number of carbon atoms implies higher stability of pore structures, which was less unaffected from excessive etching (Fig. 5G and B). The negative correlations between C content and SSA/TPV from our model are consistent with experimental results from Zhang *et al.*, Xi *et al.*, and Li *et al.*^{37,39,40}

- 15 Y. Li, R. Gupta and S. You, *Bioresour. Technol.*, 2022, **359**, 127511.
- 16 X. Yuan, M. Suvarna, S. Low, P. D. Dissanayake, K. B. Lee, J. Li, X. Wang and Y. S. Ok, *Environ. Sci. Technol.*, 2021, **55**, 11925–11936.
- 17 X. Yang, C. Yuan, S. He, D. Jiang, B. Cao and S. Wang, *Fuel*, 2023, **331**, 125718.
- 18 W. A. M. Wickramaarachchi, M. Minakshi, X. Gao, R. Dabare and K. W. Wong, *Chem. Eng. J. Adv.*, 2021, **8**, 100158.
- 19 K. M. Jablonka, D. Ongari, S. M. Moosavi and B. Smit, *Chem. Rev.*, 2020, **120**, 8066–8129.
- 20 T. Wang, R. Pan, M. L. Martins, J. Cui, Z. Huang, B. P. Thapaliya, C.-L. Do-Thanh, M. Zhou, J. Fan, Z. Yang, M. Chi, T. Kobayashi, J. Wu, E. Mamontov and S. Dai, *Nat. Commun.*, 2023, **14**, 4607.
- 21 W