Journal of Zhejiang University-SCIENCE A in press
www.jzus.zju.edu.cn; www.springer.com/journal/11582
E-mail: jzus_a@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.A2500628

Mechanism-enhanced multitask distillation for predictive, interpretable design of biomass-based activated carbons

Peng ZHAO[1], Jiahui ZHOU[2], Guangyao LI[1,3], Hao YU[1,3] ✉, Dongxu JI[3], Takahiko MIYAZAKI[1,4], Kyaw THU[1,4]

[1]Department of Advanced Environmental Science and Engineering, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, Fukuoka 816-8580, Japan
[2]Department of Information Science and Technology, Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan
[3]School of Science and Engineering, Chinese University of Hong Kong, Shenzhen 518172, China
[4]Research Center for Next Generation Refrigerant Properties (NEXT-RP), International Institute for Carbon-Neutral Energy Research (I2CNER), Kyushu University, Fukuoka, Japan

Abstract: Designing biomass-derived activated carbons (ACs) is challenged by heterogeneous synthesis routes and multiobjective trade-offs among specific surface area (SBET), total pore volume (VT), and mass yield (Yield). This study presents a mechanism-enhanced multitask distillation framework (AC-ResKD) built on a shared residual ResDNN and dual priors from prediction-level (PD-KD) and teacher-aware (TA-KD) distillation. Process factors (agent, activation temperature/time, impregnation, heating rate, precursor composition) are modeled jointly with teacher predictions to learn an interpretable mapping across combined and separated one-step/two-step datasets. All results are reported as the mean and 95% confidence intervals over 20 repeated random 80/20 splits. On separated routes, TA-KD achieves robust accuracy for SBET (one-step: R^2=0.806 [0.787, 0.824]; two-step: R^2=0.801 [0.773, 0.830]) and VT (one-step: R^2=0.787 [0.764, 0.810]; two-step: R^2=0.817 [0.791, 0.843]). On the combined set, TA-KD yields the strongest gains for Yield (R^2=0.816 [0.802, 0.831]; RMSE=5.047 [4.853, 5.242]) while improving SBET and VT as well. Overall, Yield is the most consistent beneficiary of distillation, and route-separated training exhibits improved monotonicity and reduced bias relative to combined training; the two-step route shows stronger VT predictability consistent with a pre-carbonization priming effect. Explainability (PFI/SHAP) identifies teacher outputs, agent type, and thermal severity as dominant drivers. Impregnation governs VT in one-step activation, while pyrolysis variables rise in importance in two-step activation. Robust Pareto screening with quantile-window extraction delivers agent- and route-specific operating envelopes (temperature–dose–time), enabling simultaneous SBET/VT improvement under bounded Yield penalties. AC-ResKD thus provides accurate, interpretable, and actionable guidance for AI-assisted AC design in heterogeneous, data-scarce settings.

Key words: Activated carbon synthesis, Pore structure prediction, Knowledge distillation, Multitask deep learning, Process optimization

1 Introduction

The urgency of the transition to sustainable energy systems is underscored by escalating global carbon emissions, which reached approximately 37.5 Gt (fossil CO_2), necessitating the immediate deployment of effective capture and storage technologies (Kundu et al., 2024; Lamb et al., 2022). However, this transition is constrained by intermittency, storage inefficiency, and lifecycle emissions across production chains. High-performance porous materials provide essential functions that mitigate these bottlenecks, including adsorption, separation, catalysis, and energy storage (Li et al., 2025a; Yu et al., 2023). Among these, activated carbon (AC) occupies a central role due to its unique combination of large specific surface area (S_{BET}), tunable pore-size distributions, chemical stability, and low-cost manufacture from biomass precursors (Heidarinejad et al., 2020; Sevilla and Mokaya, 2014; Yudha et al., 2024). These properties make ACs indispensable for critical technologies such as post-combustion CO_2 capture, solvent purification, supercapacitors, and metal–air batteries (Li et al., 2025b;

✉ Hao YU, yuhao@cuhk.edu.cn
Hao YU, https://orcid.org/0000-0002-9146-4462

Received Nov. 30, 2025; Revision accepted Apr. 8, 2026;
Crosschecked

Yu et al., 2024; Zhou et al., 2026). Despite the dominance of zeolites and MOFs in the academic literature, ACs remain the industrial benchmark for stability and regeneration ease (Devi et al., 2023; Kaur et al., 2023).

Activated carbon production typically follows physical activation (e.g., steam or CO_2 at high temperature) or chemical activation (e.g., KOH, H3PO4, ZnCl2) routes, implemented via either one-step (coupled carbonization–activation) or two-step (pyrolysis followed by activation) protocols (Fu et al., 2023; Othman et al., 2013). Process severity—governed by temperature, residence time, heating rate, and impregnating agent dose—determines pore development, while precursor composition (volatile matter (VM), ash, fixed carbon (FC)) sets the structural stability of the evolving carbon skeleton (Chairunnisa et al., 2024; Muhammed et al., 2021; Sosa et al., 2023). Because application performance strictly depends on a delicate balance among SBET, total pore volume (VT), and Yield, the predictive control of synthesis remains a "trial-and-error" challenge, consuming significant experimental resources.

The paradigm of materials discovery is undergoing a fundamental shift from trial-and-error experimentation to data-driven inverse design. In the broader landscape of materials science, advanced architectures such as Graph Neural Networks (GNNs) and generative transformers have achieved breakthroughs in crystalline and ordered systems (Przybek et al., 2025; Reiser et al., 2022; Zhao et al., 2025). For instance, Merchant et al. recently demonstrated the power of deep learning by discovering 2.2 million new stable inorganic crystals via the GNoME model, radically expanding the known materials horizon (Merchant et al., 2023). Similarly, in the domain of Metal-Organic Frameworks (MOFs), machine learning pipelines have been successfully deployed to screen millions of candidates for gas adsorption and catalysis by leveraging their well-defined crystallographic information files (CIFs) (Song et al., 2025). Kolbadinejad et al. successfully predicted the adsorption amounts of argon, xenon, krypton, and oxygen on activated carbon and zeolite using artificial neural networks (especially multilayer perceptrons), verifying the high efficiency of deep learning in gas adsorption simulation (Kolbadinejad et al., 2022).

However, translating these successes to Activated Carbon (AC) presents unique challenges due to its turbostratic, amorphous structure and the complex, non-linear kinetics of biomass pyrolysis, which lack simple structure-property descriptors (Belluati et al., 2024; Giudicianni et al., 2021). Consequently, AI applications in AC synthesis have primarily relied on classical regressors on tabular experimental data rather than structural graph generation (Shukla et al., 2024; Zhang et al., 2020). Researchers have employed algorithms such as Gradient Boosting Decision Trees (GBDT), Random Forests (RF), and Artificial Neural Networks (ANN) to map synthesis conditions to textural outputs (Ibrahim and Hussein, 2025; Li et al., 2024). For instance, Yuan et al. successfully employed GBDT to map biomass properties to CO_2 adsorption capacity, achieving high predictive accuracy (R2 approx 0.84 on test data) (Yuan et al., 2021). Similarly, recent work by Ibrahim et al. utilized ensemble learning to optimize texture properties (Ibrahim and Hussein, 2025).

Despite these initial successes, significant barriers prevent the widespread adoption of AI in robust AC design. First, regarding applicability and generalization, current predictive frameworks are heavily constrained by the heterogeneity of biomass precursors and the small size of experimental datasets. Most existing models, such as that of Liao et al., were often constrained by small-sample datasets (e.g., < 200 points), limiting their generalization across diverse precursors and activation agents (Liao et al., 2019). Consequently, these "local" models suffer from poor transferability. A regressor optimized for KOH-activated coconut shell often fails catastrophically when applied to H3PO4-activated lignocellulosic residues due to fundamentally different reaction kinetics and functional group evolutions.

Second, and perhaps more critically, a fundamental methodological gap exists in how target properties are modeled. The prevailing paradigm, as seen in most current studies, relies on "single-task" learning strategies: separate, independent models are trained to predict SBET, VT, and Mass Yield individually (Yuan et al., 2021). This decoupled approach ignores the intrinsic physicochemical correlations and competitive trade-offs governing carbonization—most notably, the antagonistic relationship where increasing activation severity enhances porosity and (SBET) but inevitably degrades the carbon skeleton (Yield). By treating these outputs as isolated targets, independent models often generate thermodynamically inconsistent predictions—such as forecasting ultrahigh surface areas cooccurring with high yields—which are physically unrealizable (Evans et al., 1999; Lu et al., 2017). Furthermore, standard models often function as "black boxes" that lack the capability to simultaneously optimize these conflicting objectives within a single differentiable framework. There is a scarcity of advanced Multi-Task Learning (MTL) or knowledge distillation architectures in this domain that can reconcile these trade-offs while compensating for data sparsity.

Beyond prediction accuracy, biomass activated-carbon synthesis data pose a distinct learning challenge: multiple activation mechanisms (agent identity and one-step vs. two-step routes) are mixed within a small tabular dataset, producing multimodal mappings, while SBET, VT and Yield are physically coupled and exhibit different noise/sensitivity levels. Existing MTL or KD pipelines for tabular material data often treat distillation as a generic regularizer, without explicitly addressing mechanism mixing or the need to translate multiobjective predictions into actionable operating guidance. This study addresses this gap by combining two complementary forms of teacher information—(i) prediction-level soft priors to stabilize multitarget learning under heterogeneous labels and (ii) teacher-aware proxy features to help disambiguate mixed mechanisms—together with robust Pareto screening and quantile-window extraction for decision-support window identification.

To address these limitations, AC-ResKD (Activated-Carbon Residual Knowledge Distillation) is proposed, a mechanism-aware multitask distillation framework that operationalizes the above design: prediction-level distillation and teacher-aware proxy augmentation for mechanism-mixed synthesis data, coupled with Pareto-based operating-window extraction for decision support. The approach centers on a shared ResDNN backbone and integrates dual-channel priors: prediction-level distillation (PD-KD) and teacher-aware distillation (TA-KD). Process factors ("agent—temperature/time—impregnation—heating rate—precursor composition") and high-level teacher predictions are embedded into an interpretable input–output mapping with multihead outputs for SBET, VT, and Yield. Under unified training/evaluation across combined and separated one-step/two-step datasets, the models reproduce domain-consistent monotonicity and surpass the typical benchmarks observed in the literature (Zou et al., 2024). Beyond prediction, robust Pareto screening and quantile-window extraction are implemented to provide agent- and route-specific optimal operating intervals. These windows identify conditions under which SBET and VT can be jointly increased while controlling Yield penalties, furnishing actionable guidance for process scale-up. In summary, AC-ResKD delivers a more accurate and comprehensive AI-based synthesis assistant capable of reconciling multitarget trade-offs and predicting optimal operating intervals under varied design requirements.

2 Method

2.1 Feature representation and targets

The dataset in this article is cited from Zou et al. (Zou et al., 2024), which collects synthesis conditions and textural properties of activated carbons produced with multiple chemical activating agents. Three datasets consisted of three parts: the data of activated carbons by one-step activation; the data of activated carbons by two-step activation; and the combined data from both one-step activation and two-step activation. Each record corresponds to one synthesis experiment and contains.

(i) a categorical variable indicating the activating agent (e.g., KOH, H_3PO_4, $ZnCl_2$, …),

(ii) several continuous process variables (such as precursor composition, soaking time and temperature, activation temperature and time, heating rate), and

(iii) three regression targets: specific surface area (SBET) (m^2/g), total pore volume (VT) (cm^3/g), and carbon yield (Yield) (%).

For learning, a mixed feature vector $\boldsymbol{x}$ was constructed that combines the categorical agent label and the continuous synthesis variables. The activating agent is one-hot encoded into a fixed-length binary vector. All remaining continuous variables are scaled to $[0,1]$ using a MinMaxScaler fitted on the training portion only to avoid target leakage. The final input vector is:

$$\boldsymbol{x} = [\boldsymbol{x}_{\text{agent}} \quad , \boldsymbol{x}_{\text{proc}} \quad] \in \mathbb{R}^F \tag{1}$$

where F denotes the total number of input features after one-hot expansion. The three prediction targets are denoted by:

$$y_1 = S_{\text{BET}}, y_2 = \text{VT}, y_3 = \text{Yield}. \tag{2}$$

For numerical stability, each target is also individually rescaled to $[0,1]$ on the training set before optimization. All performance metrics reported in this work (R^2, RMSE, Pearson's r) are computed after inverting this scaling back to the original physical units so that they are directly interpretable by materials scientists.

Although several activated-carbon datasets with similar variables exist in many studies (Ibrahim and Hussein, 2025; Zou et al., 2024), this work deliberately adopts this dataset as a primary test bed for two reasons. First, it is relatively challenging: the synthesis conditions are highly heterogeneous across activating agents, and the experimental points are sparsely and irregularly distributed in the process-parameter space rather than forming dense grids (Machine learning application for predicting key properties of activated carbon produced from lignocellulosic biomass waste with chemical activation). In practice, the same activating agent is often explored at only a few discrete operating points, which makes it difficult for purely data-driven models to interpolate smoothly and capture global trends. Second, the dataset is provided in a clean tabular format with well-defined physical variables, which makes it a representative benchmark for evaluating tabular deep-learning architectures under realistic data scarcity. To address robustness concerns beyond a single curated source, we further report an external evaluation on the independent dataset (Ibrahim and Hussein, 2025), which differs in variable schema and targets and therefore provides a complementary generalization check.

In addition to the original combined dataset, which merges one-step and two-step activation protocols, this work also evaluates the models on two separate subsets: a one-step dataset and a two-step dataset containing only experiments with the corresponding processing route. As shown later in Section 3.1, the combined dataset leads to systematically lower performance than training separate models on one-step and two-step data. This suggests that pooling distinct synthesis routes into a single model introduces conflicting patterns and reduces the effective smoothness of the input–output mapping. In contrast, the proposed architecture remains reasonably accurate on this difficult combined setting and performs even better on the cleaner one-step and two-step subsets, indicating that the method can be transferred to other activated-carbon datasets by only adapting the input features and the number of output heads.

Finally, the feature-importance and partial-dependence analyses in Section 3.2 reveal physically meaningful trends: the model recovers the expected influence of activation temperature, precursor composition (VM, Ash, FC), soaking conditions and agent type on SBET, VT and Yield. This consistency with domain knowledge supports that the proposed network is not only predictive but also captures plausible latent structures in the synthesis–property relationship.

2.2 Model design and structure

The model structure is shown in Figure 1, consisting of a teacher model and a student model, which are distilled through different methods. The specific method will be explained in the following sections.

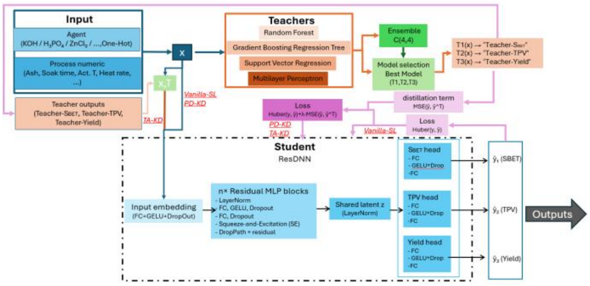


Fig. 1 Model structure

2.2.1 Teacher model library and per-target selection

As a strong nonneural reference, a heterogeneous teacher library was constructed on top of the same feature representation described in Section 2.1. The candidate teachers included tree-based, kernel-based, instance-based, neural, and simple ensemble regressors so that different inductive biases could be explored for different property targets. Rather than assuming that a single teacher family is uniformly optimal for all outputs, this work selected the teacher separately for each target. Detailed definitions of the candidate teacher models and their ensemble forms are provided in the Electronic Supplementary Materials (Section S1, Eq. (S1)–(S7)).For each outer train/test split and for each target y_k, model selection is performed only on the training set. Let $\mathcal{M}_k$ be the set of all candidate teachers for target k and denote the set of candidate teachers for target t, including both individual regressors and simple equal-weight ensembles. For any candidate model $m \in \mathcal{M}_k$ denoted, its validation coefficient of determination is computed as

$$R_k^2(m) \;=\; 1 - \frac{\sum_{i \in \mathcal{V}} (y_{k,i} - \hat{y}_{k,i}^{(m)})^2}{\sum_{i \in \mathcal{V}} (y_{k,i} - \grave{y}_k)^2} \tag{3}$$

where $\mathcal{V}$ indexes the validation samples within the training portion, $y_{k,i}$ is the observed value of target $t, \hat{y}_{k,i}^{(m)}$ is the corresponding prediction from candidate teacher m, and $\grave{y}_k$ is the mean of $y_{k,i}$ over $\mathcal{V}$. A cross-validated estimation of $R_k^2(m)$ was then used to determine the final teacher for each target:

$$m_k^* \;=\; \underset{m \in \mathcal{M}_k}{\operatorname{argmax}} R_k^2(m) \tag{4}$$

This procedure yields three potentially different best teachers, one for SBET, one for VT, and one for Yield, thereby allowing the teacher assignment to adapt to target-specific response characteristics rather than enforcing a single shared teacher across all outputs.

Once selected, each teacher was refit on the full training subset of the corresponding split and then frozen. The resulting teacher predictions were generated for both training and test samples and subsequently used in two ways in the student framework described below: (i) as prediction-level soft supervision in PD-KD and (ii) as high-level proxy features in TA-KD. Because teacher selection, refitting, and student training were all restricted to the same training portion in each split, the comparison among Vanilla-SL, PD-KD, and TA-KD remained strictly leakage-free and directly comparable under identical data partitions.

2.2.2 Student architecture (ResDNN multitask regressor)

The student model is a shared-backbone, multihead regressor operating on the same feature vector as the teacher models. A residual MLP for tabular data (ResDNN) is adopted because SBET, VT, and Yield are driven by the same synthesis variables and therefore benefit from hard parameter sharing (Gorishniy et al., 2021; Kadra et al., 2021). (Caruana, 1997; Ruder, 2017). Given an input vector $\boldsymbol{x}$, the network first applies a linear stem followed by GELU activation and dropout:

$$\boldsymbol{h}^{(\mathbf{0})} = \mathrm{Dropout}\big(\mathrm{GELU}(\boldsymbol{W}_{\mathbf{stem}}\boldsymbol{x} + \boldsymbol{b}_{\mathbf{stem}})\big) \qquad (5)$$

This is followed by preactivation residual blocks:

$$\begin{aligned}
\boldsymbol{u}^{(l)} &= \mathrm{LayerNorm}\big(\boldsymbol{h}^{(l-1)}\big),\\
\boldsymbol{v}^{(l)} &= \mathrm{Dropout}\left(\phi\left(\boldsymbol{W}_2^{(l)}\,\mathrm{Dropout}\left(\phi\left(\boldsymbol{W}_1^{(l)}\boldsymbol{u}^{(l)} + \boldsymbol{b}_1^{(l)}\right)\right) + \boldsymbol{b}_2^{(l)}\right)\right),\\
\tilde{\boldsymbol{v}}^{(l)} &= \mathrm{SE}\big(\boldsymbol{v}^{(l)}\big),\\
\boldsymbol{h}^{(l)} &= \boldsymbol{h}^{(l-1)} + \alpha_l\,\mathrm{DropPath}\big(\tilde{\boldsymbol{v}}^{(l)}\big),
\end{aligned} \qquad (6)$$

where GELU denotes the activation function, $\boldsymbol{W}_{\mathbf{1}}^{(l)}$ and $\boldsymbol{W}_{\mathbf{2}}^{(l)}$ are the block parameters, and SE/DropPath denote feature reweighting and stochastic depth, respectively. After the last block, a shared latent representation is obtained:

$$\boldsymbol{z} = \mathrm{LayerNorm}\big(\boldsymbol{h}^{(L)}\big) \in \mathbb{R}^{d_{\mathrm{model}}} \qquad (7)$$

Three task-specific regression heads are then applied to predict the scaled targets for SBET, VT, and Yield:

$$\hat{y}_k = f_k(z) = \boldsymbol{w}_{\boldsymbol{k},\mathbf{2}}^{\mathrm{T}}\,\phi\big(\boldsymbol{W}_{\boldsymbol{k},\mathbf{1}}\boldsymbol{z} + \boldsymbol{b}_{\boldsymbol{k},\mathbf{1}}\big) + b_{k,2}, k = 1,2,3 \quad (8)$$

All heads are optimized jointly through a summed regression loss. Regularization is implemented using dropout and stochastic depth, and hyperparameters are selected on the training set only.

2.2.3 Knowledge distillation variants

Three student-training paradigms were compared using the same ResDNN backbone. Let $\hat{y}_k(\mathbf{x})$ and $\hat{y}_k^{(\mathrm{T})}(\mathbf{x})$ denote the student and teacher predictions for target y_k, respectively. All losses are computed on individually scaled targets using the Huber loss. The student is trained only on the ground-truth labels:

(1) Vanilla supervised learning (Vanilla-SL).

$$L_{\mathrm{Vanilla}} = \sum_{k=1}^{3} w_k\, l_\delta(\hat{y}_k(\boldsymbol{x}), y_k) \qquad (9)$$

where w_k are task weights. This serves as the deep learning baseline.

(2) Prediction-level knowledge distillation (PD-KD).

A distillation term is added to encourage agreement with the selected teacher:

$$L_{\mathrm{PD\text{-}KD}} = \sum_{k=1}^{3} w_k\big[l_\delta(\hat{y}_k(\boldsymbol{x}), y_k) \;+\; \lambda_{\mathrm{KD}}\cdot\| \hat{y}_k(\boldsymbol{x}) - \hat{y}_k^{(\mathrm{T})}(\boldsymbol{x})\|_2^2\big] \quad (10)$$

where λ_{KD} is a scalar trade-off coefficient.

(3) Teacher-aware knowledge distillation (TA-KD)

The teacher predictions are further concatenated with the original input as auxiliary features,

$$\tilde{\boldsymbol{x}} = \big[\boldsymbol{x}, \hat{y}_1^{(\mathrm{T})}, \hat{y}_2^{(\mathrm{T})}, \hat{y}_3^{(\mathrm{T})}\big] \qquad (11)$$

and the student is optimized with the same supervised–distillation objective:

$$L_{\mathrm{TA\text{-}KD}} = \sum_{k=1}^{3} w_k\big[l_\delta(\hat{y}_k(\tilde{\boldsymbol{x}}), y_k) \;+\; \lambda_{\mathrm{KD}}\cdot\| \hat{y}_k(\tilde{\boldsymbol{x}}) - \hat{y}_k^{(\mathrm{T})}(\boldsymbol{x})\|_2^2\big] \quad (12)$$

In all experiments, the distillation weight was fixed to $\lambda_{\mathrm{KD}} = 0.5$ for PD-KD and $\lambda_{\mathrm{KD}} = 0.7$ for TA-KD. These values were chosen empirically to provide a balanced teacher signal without overwhelming the supervised loss and were kept unchanged across all targets, datasets, and repeated splits for consistency.

2.2.4 Training and evaluation protocol

For each dataset variant (Combined, One-step, and Two-step), model performance was assessed using 20 repeated random 80/20 train/test splits. In each repeat, 80% of the samples were used for model development, and the remaining 20% were held out for final evaluation. The same set of repeated splits was reused across all learning paradigms (Vanilla-SL, PD-KD, TA-KD, and Teacher) to enable paired and directly comparable performance assessment. All reported R^2, RMSE, and Pearson's r values were computed on the held-out test subsets after inverse transformation to the original physical units and are summarized as the mean values with 95% confidence intervals over the 20 repeats.

Within each training portion, all preprocessing and model-selection steps were performed without access to the test subset. Specifically, categorical encoding and Min–Max scaling were fitted using the training data only, and each target was scaled individually for optimization before being mapped back to physical units for evaluation. The teacher model for each target was selected only from the training portion, as described in Section 2.2.1, and then refitted on the full training subset of the corresponding split. A fresh ResDNN student was subsequently initialized for each paradigm and trained under the same optimization protocol, differing only in whether teacher information was used in the loss term (PD-KD) and/or as auxiliary input features (TA-KD). This design ensures that performance differences among the student variants can be attributed to the use of teacher information rather than to differences in data partitioning, preprocessing, or training budget.

To examine robustness beyond the primary dataset, the same evaluation pipeline was also applied to an independent activated-carbon dataset with different descriptors and targets. The corresponding benchmark results are reported in the Electronic Supplementary Materials (Section S2, Table S1).

2.3 Robust Pareto screening and optimal-interval extraction

To translate the trained surrogate models into actionable synthesis guidance, a robust Pareto screening procedure followed by quantile-window extraction was performed for the most frequent activating agents. For each dataset and agent, a feasible candidate domain was constructed by restricting each continuous process variable to the empirical 2nd–98th percentile range observed in the training data while fixing categorical variables to the target agent (and dominant category where applicable). Monte Carlo candidate points were then generated:

$$\mathcal{X}_a = \{\boldsymbol{x}_{\boldsymbol{a}}^{(\boldsymbol{i})}\}_{i=1}^{N} \qquad (13)$$

and transformed using the same preprocessing pipeline as in model training.

Each candidate was evaluated by two student models, Vanilla-SL and PD-KD, yielding predictions for SBET, VT, and Yield:

$$M_k\colon \widetilde{\boldsymbol{x}}_{\boldsymbol{a}}^{(\boldsymbol{i})} \mapsto (\hat{y}_{k,1}^{(i)}, \hat{y}_{k,2}^{(i)}, \hat{y}_{k,3}^{(i)}) = (\widehat{S_{\mathrm{BET}}}, \widehat{\mathrm{VT}}, \widehat{\mathrm{Yield}})_k^{(i)} \quad (14)$$

These outputs were combined into a six-dimensional objective vector,

$$\boldsymbol{f}(\boldsymbol{x}_{\boldsymbol{a}}^{(\boldsymbol{i})}) = (\hat{y}_{1,1}^{(i)}, \hat{y}_{1,2}^{(i)}, \hat{y}_{1,3}^{(i)}, \hat{y}_{2,1}^{(i)}, \hat{y}_{2,2}^{(i)}, \hat{y}_{2,3}^{(i)})^{\top} \in \mathbb{R}^6 \# (15)$$

with all objectives treated as larger-is-better. A candidate was considered dominant if it was no worse in all objectives and strictly better in at least one:

$$\boldsymbol{f}(\boldsymbol{x}_{\boldsymbol{a}}) \succ \boldsymbol{f}(\boldsymbol{z}_{\boldsymbol{a}}) \Leftrightarrow f_m(\boldsymbol{x}_{\boldsymbol{a}}) \geq f_m(\boldsymbol{z}_{\boldsymbol{a}})\ \forall m \text{ and } f_m(\boldsymbol{x}_{\boldsymbol{a}}) > f_m(\boldsymbol{z}_{\boldsymbol{a}}) \text{ for at least one } m \qquad (16)$$

The robust Pareto set for each agent was then defined as the nondominated subset:

$$\mathcal{P}_a^{\mathrm{robust}} = \left\{\boldsymbol{x}_{\boldsymbol{a}}^{(\boldsymbol{i})} \in \mathcal{X}_a \middle| \nexists\, \boldsymbol{z}_{\boldsymbol{a}} \in \mathcal{X}_a \text{ s.t. } \boldsymbol{f}(\boldsymbol{z}_{\boldsymbol{a}}) \succ \boldsymbol{f}\left(\boldsymbol{x}_{\boldsymbol{a}}^{(\boldsymbol{i})}\right)\right\} \quad (17)$$

When necessary, quasi-Pareto fallback and crowding-distance subsampling were used to maintain numerical stability and diversity.

To obtain practically usable operating ranges, the Pareto set was summarized by narrow quantile windows. For each continuous variable, the 45th and 55th percentiles over the Pareto set were extracted:

$$[L_j^{\star}(a), U_j^{\star}(a)] = [q_{45\%}(x_j|\boldsymbol{x} \in \mathcal{P}_a^{\mathrm{robust}}),\ q_{55\%}(x_j|\boldsymbol{x} \in \mathcal{P}_a^{\mathrm{robust}})] \qquad (18)$$

These quantile intervals define Pareto-consistent operating windows in which SBET, VT, and Yield remain jointly favorable and predictions from different student models are mutually consistent. The resulting ranges are reported in Section 3.3 as agent- and route-specific best-interval recommendations.

3 Results and discussion

3.1 Comparison of model results

Table 1 The model results of different datasets

Dataset	Method	Target	***R*²** **(mean [95%CI])**	**Pearson** **(mean [95%CI])**	**RMSE** **(mean [95%CI])**
Combined	Vanilla-SL	SBET (m2/g)	0.708 [0.679, 0.736]	0.847 [0.831, 0.864]	346.684 [328.088, 365.279]
		VT (cm3/g)	0.658 [0.624, 0.693]	0.816 [0.796, 0.837]	0.203 [0.192, 0.213]
		Yield (%)	0.577	0.771	7.646

			[0.538, 0.616]	[0.747, 0.794]	[7.325, 7.968]
	PD-KD	SBET (m2/g)	0.751 [0.728, 0.775]	0.871 [0.858, 0.884]	319.779 [303.708, 335.851]
		VT (cm3/g)	0.716 [0.690, 0.742]	0.853 [0.839, 0.866]	0.185 [0.176, 0.194]
		Yield (%)	0.744 [0.719, 0.770]	0.869 [0.855, 0.882]	5.937 [5.651, 6.223]
	TA-KD	SBET (m2/g)	0.768 [0.751, 0.784]	0.882 [0.873, 0.890]	309.643 [299.841, 319.446]
		VT (cm3/g)	0.739 [0.720, 0.757]	0.866 [0.855, 0.877]	0.177 [0.170, 0.185]
		Yield (%)	0.816 [0.802, 0.831]	0.909 [0.899, 0.918]	5.047 [4.853, 5.242]
	Teacher	SBET (m2/g)	0.760 [0.746, 0.774]	0.880 [0.872, 0.889]	315.384 [304.881, 325.887]
		VT (cm3/g)	0.725 [0.710, 0.741]	0.859 [0.849, 0.870]	0.183 [0.174, 0.191]
		Yield (%)	0.809 [0.796, 0.821]	0.910 [0.901, 0.919]	5.164 [4.953, 5.375]
One-step	Vanilla-SL	SBET (m2/g)	0.758 [0.739, 0.776]	0.878 [0.867, 0.889]	322.039 [307.767, 336.312]
		VT (cm3/g)	0.737 [0.717, 0.757]	0.869 [0.857, 0.881]	0.177 [0.170, 0.185]
		Yield (%)	0.705 [0.672, 0.737]	0.850 [0.832, 0.869]	6.263 [5.791, 6.735]
	PD-KD	SBET (m2/g)	0.796 [0.778, 0.814]	0.900 [0.889, 0.910]	294.821 [280.225, 309.417]
		VT (cm3/g)	0.785 [0.760, 0.810]	0.896 [0.882, 0.910]	0.159 [0.152, 0.167]
		Yield (%)	0.782 [0.754, 0.810]	0.894 [0.879, 0.909]	5.376 [4.933, 5.819]
	TA-KD	SBET (m2/g)	0.806 [0.787, 0.824]	0.906 [0.896, 0.915]	286.991 [274.933, 299.050]
		VT (cm3/g)	0.787 [0.764, 0.810]	0.896 [0.884, 0.908]	0.159 [0.151, 0.167]
		Yield (%)	0.825 [0.807, 0.843]	0.918 [0.908, 0.928]	4.815 [4.511, 5.119]
	Teacher	SBET (m2/g)	0.805 [0.790, 0.820]	0.903 [0.894, 0.911]	288.580 [276.628, 300.533]
		VT (cm3/g)	0.786 [0.769, 0.802]	0.893 [0.883, 0.903]	0.160 [0.153, 0.168]
		Yield (%)	0.835 [0.813, 0.856]	0.923 [0.912, 0.934]	4.662 [4.313, 5.012]
Two-step	Vanilla-SL	SBET (m2/g)	0.746 [0.702, 0.791]	0.872 [0.847, 0.897]	268.615 [240.862, 296.367]
		VT (cm3/g)	0.782 [0.733, 0.830]	0.894 [0.869, 0.918]	0.138 [0.124, 0.153]
		Yield (%)	0.734 [0.708, 0.760]	0.882 [0.863, 0.901]	6.465 [5.863, 7.066]
	PD-KD	SBET (m2/g)	0.781 [0.749, 0.812]	0.893 [0.874, 0.911]	250.140 [230.457, 269.824]
		VT (cm3/g)	0.790 [0.735, 0.845]	0.902 [0.878, 0.926]	0.135 [0.120, 0.150]
		Yield (%)	0.754 [0.730, 0.778]	0.889 [0.870, 0.909]	6.215 [5.676, 6.755]

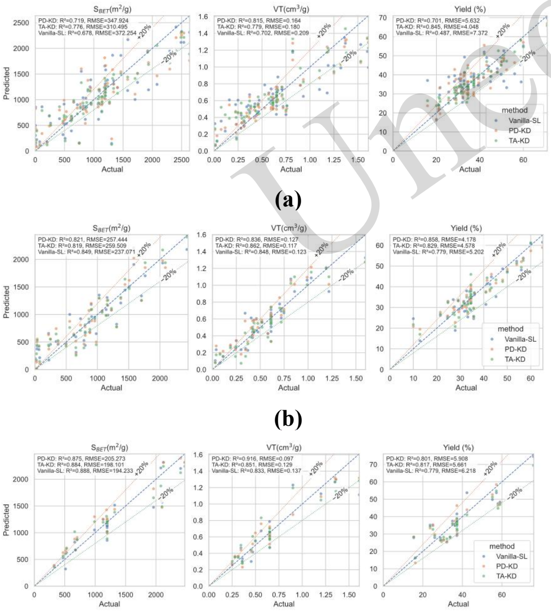
TA-KD	SBET (m2/g)	0.801 [0.773, 0.830]	0.905 [0.890, 0.920]	238.235 [218.637, 257.832]
	VT (cm3/g)	0.817 [0.791, 0.843]	0.915 [0.901, 0.930]	0.130 [0.119, 0.141]
	Yield (%)	0.791 [0.770, 0.811]	0.911 [0.895, 0.927]	5.766 [5.179, 6.353]
Teacher	SBET (m2/g)	0.785 [0.763, 0.807]	0.899 [0.887, 0.911]	249.900 [233.873, 265.927]
	VT (cm3/g)	0.805 [0.779, 0.832]	0.910 [0.894, 0.926]	0.135 [0.121, 0.148]
	Yield (%)	0.861 [0.844, 0.878]	0.936 [0.927, 0.945]	4.643 [4.215, 5.071]

This study investigates the chemical activation of biomass-derived activated carbons and examines how key process variables—activating agent type, activation temperature and duration, impregnation conditions, heating rate, and precursor composition (volatile matter, VM; ash; fixed carbon, FC)—jointly affect specific surface area (SBET), total pore volume (VT), and mass yield (Yield). A multitask residual deep network (ResDNN) together with knowledge distillation variants (prediction-level distillation, PD-KD; teacher-aware distillation, TA-KD) is adopted, and findings are cross-validated against mechanistic expectations and model behavior. Across dataset configurations (combined one-/two-step activation; separated one-step and two-step routes), the model exhibits consistent alignment with expected physical trends, captures the directionality and strength of variable–property relationships, and reflects interpretable trade-offs among the three targets.

Relative to the dataset-agnostic reference study of (Zou et al., 2024) (SBET: R^2=0.78, RMSE=293.76; VT: R^2=0.76, RMSE=0.19; Yield: R^2=0.91, RMSE=4.49), our repeated-split evaluation shows that route-separated modeling achieves competitive—and for SBET/VT often higher—accuracy in original physical units. For example, one-step TA-KD attains SBET R^2=0.806 [0.787, 0.824] and VT R^2=0.787 [0.764, 0.810], while two-step TA-KD attains SBET R^2=0.801 [0.773, 0.830] and VT R^2=0.817 [0.791, 0.843], with VT RMSE=0.130 [0.119, 0.141] below the reference RMSE of 0.19. The combined dataset remains more challenging due to stronger mechanism heterogeneity; nevertheless, distillation still yields substantial improvements (e.g., combined TA-KD: SBET R^2=0.768 [0.751, 0.784]; VT R^2=0.739 [0.720, 0.757]). For Yield, our student models do not reach the R^2=0.91 reported in (Zou et al., 2024) and exhibit higher RMSE (e.g., combined TA-KD RMSE=5.047 [4.853, 5.242]), highlighting the sensitivity of Yield to heterogeneous reporting and the fact that cross-study comparisons can be affected by differences in preprocessing choices (e.g., target scaling/inverse-scaling, split protocol, and any filtering rules) that are not fully specified in the reference. Accordingly, we treat (Zou et al., 2024) as a useful reference point rather than a strictly controlled baseline, and we report all results under a fully specified and repeatable protocol (20 repeated splits with mean and 95% CI in original units). Yield is the most consistently improved target under distillation. In the combined dataset, TA-KD increases R^2 from 0.577 [0.538, 0.616] (Vanilla-SL) to 0.816 [0.802, 0.831] and reduces RMSE from 7.646 [7.325, 7.968] to 5.047 [4.853, 5.242], delivering the strongest student performance. In the one-step dataset, TA-KD is the best-performing student for Yield (R^2=0.825 [0.807, 0.843]; RMSE=4.815 [4.511, 5.119]) and approaches the teacher baseline (R^2=0.835 [0.813, 0.856]). In the two-step dataset, TA-KD again leads among students for Yield (R^2=0.791 [0.770, 0.811]), while the teacher remains substantially stronger (R^2=0.861 [0.844, 0.878]), indicating that Yield is particularly sensitive to route-specific reporting heterogeneity and benefits from stronger inductive bias. For VT, distillation improves performance across all settings, with the most reliable gains under route-separated training: one-step TA-KD reaches R^2=0.787 [0.764, 0.810] (RMSE=0.159 [0.151, 0.167]), two-step TA-KD reaches R^2=0.817 [0.791, 0.843] (RMSE=0.130 [0.119, 0.141]), and the combined dataset remains more challenging (best student VT R^2=0.739 [0.720, 0.757]). For SBET, TA-KD provides the best or near-best student performance across dataset variants (combined R^2=0.768 [0.751, 0.784]; one-step R^2=0.806 [0.787, 0.824]; two-step R^2=0.801 [0.773, 0.830]). SBET is more sensitive to microstructural detail and extreme values; compared with PD-KD, explicit conditioning in TA-KD is more robust under combined heterogeneity.

The influence of route differences (one-step vs two-step) and data heterogeneity is corroborated by aggregated metrics and calibration behavior under repeated-split evaluation. In homogeneous routes (one-step, two-step), input–output mappings are nearly monotonic and better calibrated, and distillation consistently enhances VT and Yield. The combined dataset exhibits stronger heterogeneity; TA-KD provides the largest gains in this setting, consistent with the need for teacher-conditioned signals to mitigate multimechanism mismatches. Evidence for a precarbonization priming effect is most clearly reflected in VT: the two-step route achieves higher VT predictability (TA-KD R^2=0.817 [0.791, 0.843]) and lower VT RMSE than the one-step route, consistent with pyrolysis presetting the carbon skeleton and pore-size spectrum. The SBET performance is comparable between the one-step and two-step methods within the confidence intervals, suggesting that the priming effect manifests more strongly in pore volume formation than in surface area extremes.

To assess cross-source robustness, the framework was further tested on an independent activated-carbon dataset with different descriptors and targets. The distillation variants preserved the same performance ordering, and TA-KD remained the strongest student model(TableS1).



(a)

(b)

(c)

Fig. 2 Parity plot of three datasets. (a) Parity plots by target on dataset *combined*. (b) Parity plots by target on dataset *one-step*. (c) Parity plots by target on dataset *two-step*. For clarity, the plots visualize one illustrative split; aggregated performance over 20 repeated splits (mean and 95% CI in original units) is reported in Table X and used for all quantitative claims.

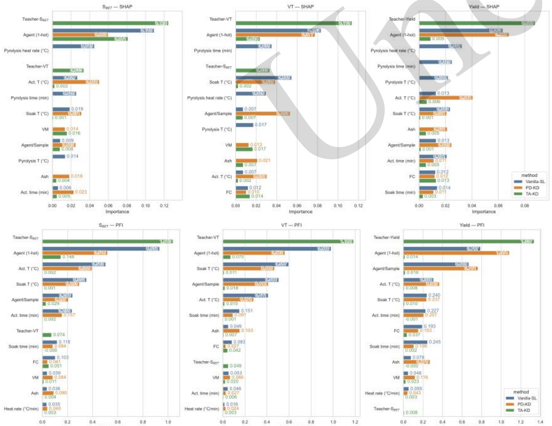
Figure 2 presents a reanalysis and comparative assessment of three parity plots, where each point represents the relationship between the mean predicted value and the mean observed value within an equal-frequency bin of predicted scores. The per-panel metrics shown in the plots correspond to the illustrated split and are not directly comparable to the aggregated statistics in Table 1. Proximity to the dashed line (y=x) indicates minimal bias and a more faithful monotonic mapping within that prediction interval. Systematic overestimation or underestimation appears as entire segments lying above or below the diagonal, respectively. Slope contraction and fan-shaped dispersion at the extremes signify regression-to-the-mean and heteroscedasticity.

The parity plots indicate that the one-step and two-step curves adhere more closely to the ideal diagonal, reflecting that a more homogeneous kinetic context facilitates learning near-linear, monotonic input–output mappings. Distillation yields clear gains for VT and Yield, with target-dependent benefits. In contrast, the combined dataset exhibits widespread slope contraction and high-end compression, attributable to the nonequivalence of route semantics, agent–operating-window covariation, unmodeled interactions between activation severity and composition, and compounded measurement heterogeneity. TA-KD substantially corrects bias for SBET and Yield, whereas PD-KD performs best for VT. Full mitigation of these structural challenges likely requires mechanistic severity features and explicit interaction terms on the feature side together with conditional or routed architectures on the model side. Following these findings, target- and route-specific distillation and post hoc calibration can render the three targets closer to y=x across the full range, yield more uniform residuals, and improve generalization under extreme operating conditions and across activating agents.

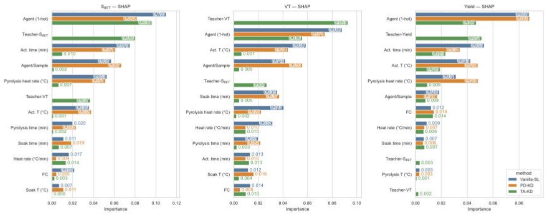
Coupling and process trade-offs among the three targets follow expected physicochemical patterns. The SBET and VT are generally positively correlated but show a characteristic rise–peak–saturation or decline. The Yield is negatively correlated with both, reflecting the cost of deeper pore development. The updated analyses confirm that a higher activation temperature or longer duration, as well as more thorough impregnation, increases SBET and VT but reduces Yield. The position on this trade-off depends on the activating agent: KOH induces stronger etching (SBET/VT increase, Yield decreases); H3PO4 promotes mid-temperature crosslinking that better preserves Yield; and ZnCl2 favors mesoporous development. Multitask parameter sharing helps the model learn these endogenous trade-offs and explains the consistent advantage of Vanilla-SL for SBET on the separated datasets. Distillation, through smoothing and conditioning, further reduces variance and bias for VT and Yield. Similar route- and target-dependent bias patterns were observed in the calibration analysis (Fig. S1), further supporting the superior monotonicity of route-separated training.

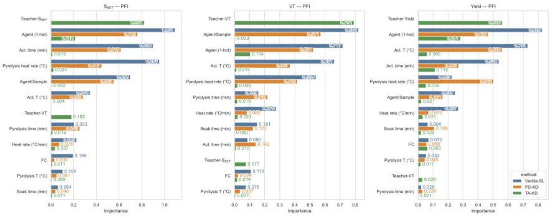
Consistent with the physicochemical mechanisms of biomass activation, the AI models reproduce domain-aligned trends for SBET, VT, and Yield. The multitask residual student, conditioned on teacher information, captures nonlinearities and interactions robustly in heterogeneous, data-scarce tabular settings. Under a unified training and evaluation protocol, the performance metrics exceed the upper bounds reported by the reference study, with especially strong gains for VT and Yield. Distillation gains for SBET are less stable on separated routes, motivating target- and route-specific distillation strengths and structural constraints. These results provide quantitative tools and physical grounding for process-parameter optimization and offer practical strategies for model transfer to broader datasets.

3.2 Model Interpretability Analysis



(a)





(b)

Fig. 3 Variable importance under the *one-step* and *two-step* modeling frameworks. (a) PFI and SHAP under dataset *one-step;* (b) PFI and SHAP under dataset *two-step*

Figure 3 illustrates the calibration of variable importance under the one-step and two-step modeling frameworks and presents a unified interpretation of how process parameters affect the predicted outputs. Two complementary explainability metrics are employed to assess feature importance: permutation feature importance (PFI), which quantifies global performance contribution by the drop in test R^2 (ΔR^2) upon shuffling a single feature, and global mean |SHAP| values, which capture the magnitude and direction of local, samplewise contributions. As comparative baselines, three student training paradigms are considered: a purely supervised multitask residual network (Vanilla-SL), prediction-level distillation (PD-KD), and teacher-aware distillation (TA-KD). Analyses span both one-step and two-step activation datasets and are cross-validated against other results in the document (combined-set behavior, calibration and parity plots) to corroborate variable ranking and physical consistency.

The importance profiles reveal consistent global regularities. Teacher outputs are the most influential features. Teacher- SBET/VT/Yield produces the largest ΔR^2 in PFI and ranks among the top contributors in SHAP, indicating that teacher predictions serve not merely as loss-level corrections but also as high-information priors actively used by the student models. Secondary dominant factors comprise the activating agent and the key thermal processing parameters. Activating agent type, activation temperature, and activation time consistently rank near the top across both datasets. In the two-step setting, the pyrolysis heating rate/temperature/duration also contribute substantially. Distillation modifies the balance rather than the ordering of feature importance. Relative to Vanilla-SL, PD-KD and TA-KD reduce overreliance on a single teacher feature and elevate the weights of process variables without altering the underlying physical conclusion regarding which variables are most important.

For the two-step dataset, SBET exhibits PFI with

Teacher-SBET as an unequivocal top feature, followed by the activating agent, activation time, pyrolysis heating rate, and sample/formulation (Agent/Sample) as a second tier. SHAP indicates that local variation is predominantly driven by agent choice and time/heating-rate parameters, with Teacher-SBET refining global bias. The SBET is governed by a balance between micropore development and pore-wall stability. The pyrolysis stage preconditions the skeleton's etchability, thereby elevating the importance of pyrolysis variables in two-step activation. For VT (total pore volume), Teacher-VT ranks first in both PFI and SHAP, followed by Agent/Sample and the activating agent, with activation temperature/time and pyrolysis rate/time forming the second tier. This pattern indicates that the total pore volume is strongly influenced by the agent chemistry and the time scales of impregnation/pyrolysis release. In two-step activation, precarbonization-induced meso-/macrostructural features (e.g., cracks) also contribute to VT. For Yield, Teacher-Yield remains first in PFI but is comparable to the activating agent in SHAP, highlighting activation temperature, pyrolysis rate, and sample/formulation as important drivers. Increasing activation temperature and duration intensifies activation severity and lowers yield. Strong bases such as KOH typically elevate SBET/VT at the expense of Yield, making the agent itself a switch for performance–yield trade-offs.

In the one-step dataset, SBET shows Teacher-SBET and agent as leading features, with activation temperature third and soaking temperature/time together with agent/sample comprising the second tier. SHAP confirms the dominant roles of Teacher-SBET and agent. This behavior reflects the absence of a dedicated pyrolysis stage in one-step activation. The SBET is primarily controlled by agent selection and the peak activation temperature, while soaking conditions regulate agent penetration uniformity and depth as secondary but nonnegligible levers. For VT, Teacher-VT is a clear first, followed by the activating agent and soaking temperature/time, with activation temperature ranked next. This indicates heightened sensitivity of VT to the impregnation stage in one-step activation. The volumetric distribution and uniformity of the activating agent predominantly determine the accessible pore volume, while the downstream activation temperature acts as an amplifier. In contrast, for Yield, the activating agent often ties or slightly exceeds Teacher-Yield in SHAP, whereas Teacher-Yield retains the top spot in PFI. The activation temperature/time and soaking temperature/time contribute strongly. Yield variation reflects the cost of deeper pore development: more thorough impregnation and more severe activation typically raise SBET/VT while lowering Yield. The models project this trade-off onto temperature/time and agent choice.

These results support a physically interpretable synthesis narrative and offer practical guidance for process design. The activating agent acts as the "steering wheel," while temperature and time serve as the "throttle." Agent chemistry (strong base, acid, salt) determines the micro/mesopore spectrum and the dominant etching pathway. The activation temperature and duration then tune the severity and enable a controllable trade-off among the SBET, VT, and Yield. The impregnation stage sets both dose and uniformity, which is especially critical for VT in one-step activation. A higher soaking temperature or longer soaking time generally increases the accessible pore volume but can also exacerbate downstream skeleton loss, thereby reducing Yield. In the two-step activation, a clear pyrolysis "priming" effect is observed. The heating rate, temperature, and time shape the initial skeleton and crack network, presetting the upper bound and stability of subsequent pore opening. These factors exert indirect yet substantial impacts on SBET, VT, and Yield. Distillation should be tailored to the target. TA-KD is preferable for structural properties (SBET, VT) under heterogeneous conditions because it strengthens conditioning. PD-KD is preferable for macroscopic targets (Yield) because it suppresses interval bias. For microstructure-sensitive SBET—particularly on separate datasets—a weaker distillation strength or selective disabling of teacher features may be warranted.

Cross-consistency is evident across importance rankings, calibration/parity plots, and performance metrics. Two-step activation exhibits more linear and monotonic behavior. In one-step activation, SBET is more sensitive to teacher smoothing, and TA-KD slightly overestimates at lower quantiles. Under two-step activation with PD-KD, VT exhibits near-perfect calibration. For Yield, PD-KD is least biased in one-step, whereas TA-KD provides the most balanced calibration in two-step. These patterns align with the test R^2/RMSE rankings: in one-step, VT is best with TA-KD, Yield with PD-KD, and SBET with Vanilla-SL; in two-step, VT is best with PD-KD, Yield with TA-KD, and SBET is more stable with Vanilla-SL/TA-KD.

Overall, the analysis indicates that teacher outputs, agent type, and key thermal parameters jointly dominate the input–output mapping for all three targets. Distillation improves robustness by redistributing information sources without altering the fundamental physical relationships. The agreement among importance ordering, calibration/parity alignment, and quantitative performance supports that the models achieve high predictive accuracy while remaining consistent with an ‘agent–activation–pore structure–yield’ synthesis logic. For process optimization, priority should be given to tuning agent selection, impregnation conditions, and activation temperature/time—and, in two-step activation, pyrolysis heating rate/temperature—supplemented by target-dependent distillation and mechanistic severity descriptors to achieve a quantifiable and interpretable balance among SBET, VT, and Yield.

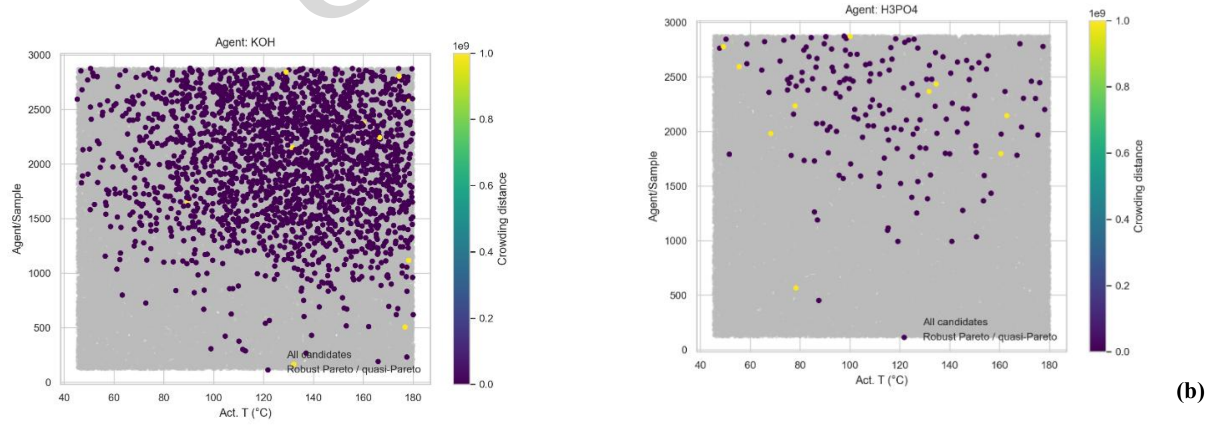
Importantly, this interpretability narrative is supported by cross-evidence stability rather than a single post-hoc view. The same small set of dominant drivers (teacher signals, route/agent descriptors, and thermal-severity variables) underpins the models that achieve the best averaged predictive performance across repeated train/test splits (Table 1). Moreover, on an independent external dataset with a different descriptor schema and targets (SA/TPV/MPV), the distillation variants preserve the same performance ordering and yield consistent improvements (Table 2), suggesting that the learned attributions reflect transferable synthesis regularities rather than split-specific artifacts.

3.3 Prediction of best interval

3.3.1 The prediction of One-step activation

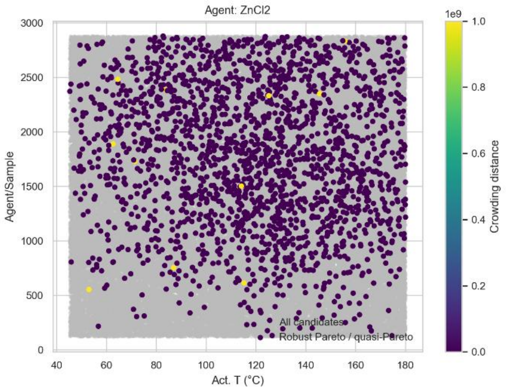
Table 3. Condensed optimal operating windows for one-step activation

Agent	Impregnation window	Activation window	Main process signature
H3PO4	Agent/sample: **4.93–5.50**; Soaking time: **2100–2260 min**; Soaking temperature: **37–40 °C**	Activation temperature: **638–682 °C**; Activation time: **109–118 min**	High dose and thorough impregnation combined with relatively high activation temperature, favoring balanced SBET–VT enhancement with controlled Yield penalty
KOH	Agent/sample: **2.80–3.21**; Soaking time: **2017–2152 min**; Soaking temperature: **41–46 °C**	Activation temperature: **452–480 °C**; Activation time: **124–133 min**	Efficient low-to-mid-temperature etching, consistent with strong-base activation, but more sensitive to Yield loss at excessive severity
ZnCl2	Agent/sample: **1.76–2.14**; Soaking time: **1725–1912 min**; Soaking temperature: **58–64 °C**	Activation temperature: **521–599 °C**; Activation time: **114–124 min**	More VT-oriented operating envelope, with soaking temperature playing a stronger role than extreme dosage



(a)

(b)



(c)

Fig. 4 Robust Pareto screening in the activation–temperature design space for each agent. (a) Agent: KOH. (b) Agent: H_3PO_4. (c) Agent: $ZnCl_2$.

Under the one-step configuration, robust Pareto screening identified clear agent-specific operating envelopes that remained favorable across the coupled trade-offs among SBET, VT, and Yield. Figure 4 visualizes the Pareto-preferred regions in the activation-temperature design space, while the condensed operating windows are summarized in Table 3. Compared with the original full-variable table, Table 3 emphasizes the most directly controllable process parameters so that the recommended operating ranges can be interpreted more intuitively at the process-design level. Full variable bounds, including feedstock-composition constraints and additional auxiliary variables, are provided in Table S2.

For H_3PO_4, the preferred one-step window is characterized by a relatively high agent dose, long soaking duration, and moderate soaking temperature, followed by activation at comparatively high temperature with moderate residence time. This pattern indicates that in the coupled one-step environment, homogeneous upstream impregnation and sufficiently strong downstream activation jointly promote balanced improvement in SBET and VT while avoiding excessive Yield loss. The Pareto band in Figure 4 is therefore consistent with a "high-dose, thorough-impregnation, moderate-duration activation" strategy.

For KOH, the optimal window shifts toward a lower activation-temperature regime with a slightly longer activation time, despite retaining a substantial impregnation requirement. This agrees with the known strong-base etching behavior of KOH, which can efficiently develop microporosity at lower thermal loads than the acidic route. However, the corresponding Pareto region also indicates a stronger sensitivity of Yield to excessive activation severity, implying that the KOH window should be controlled more carefully to prevent overetching.

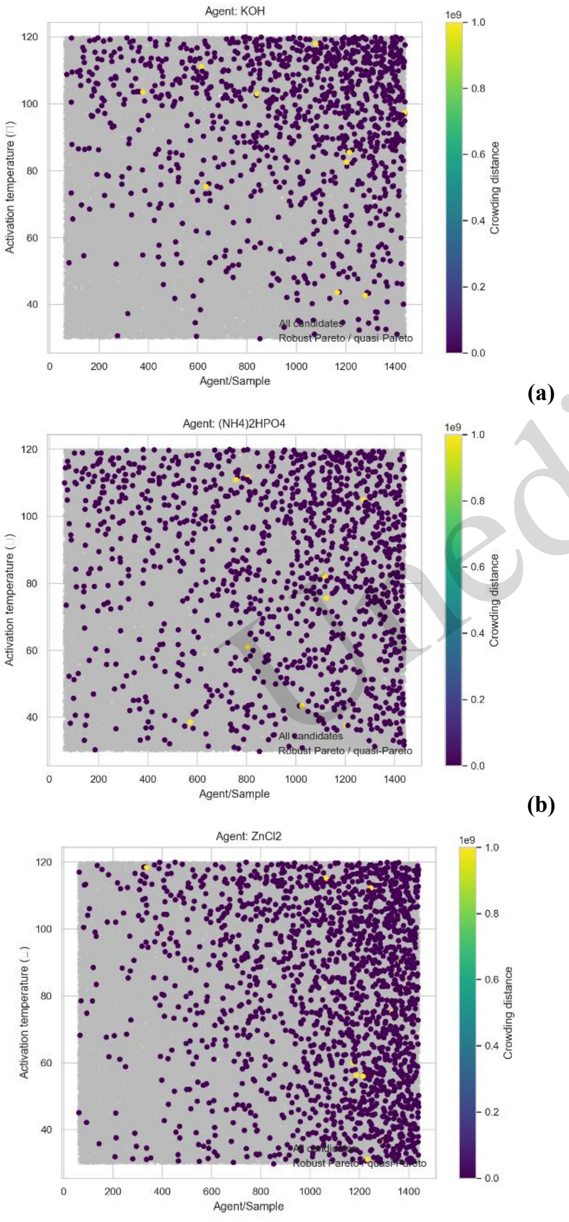
For $ZnCl_2$, the preferred envelope is comparatively less dose-intensive but more dependent on elevated soaking temperature, suggesting that solute mobility and impregnation effectiveness play an important role in the resulting pore-volume development. In this route, the activation window remains in a mid-to-high temperature range with moderate duration, and the resulting Pareto region is more clearly aligned with stable VT enhancement than with extreme SBET maximization. Overall, the one-step results reveal distinct agent-specific operating signatures rather than a single universal optimum, highlighting that the recommended process window must be conditioned on activation chemistry.

3.3.2. The prediction of Two-step activation

Table 4. Condensed optimal operating windows for two-step activation

Agent	Impregnation window	Pyrolysis window	Activation window	Main process signature
H3PO4	Agent/sample: **1.21–1.60**; Soaking time: **979–1106 min**; Soaking temperature: **68–79 °C**	Pyrolysis temperature: **698–786 °C**; Pyrolysis time: **162–204 min**	Activation temperature: **786–835 °C**; Activation time: **86–95 min**	Pyrolysis preforming followed by high-temperature but controlled-duration activation, supporting concurrent SBET and VT development
KOH	Agent/sample: **1.85–2.42**; Soaking time: **1023–1128 min**; Soaking temperature: **76.5–84.9 °C**	Pyrolysis temperature: **650–743 °C**; Pyrolysis time: **306–353 min**	Activation temperature: **633–715 °C**; Activation time: **100–105 min**	Stronger precarbonization reduces the need for prolonged downstream etching and shifts the optimum toward shorter activation
ZnCl2	Agent/sample: **0.78–0.96**;	Pyrolysis	Activation	A stable VT-oriented compromise

Agent	Impregnation window	Pyrolysis window	Activation window	Main process signature
	Soaking time: **1115–1191 min**; Soaking temperature: **68.6–77.8 °C**	temperature: **746–807 °C**; Pyrolysis time: **188–236 min**	temperature: **799–832 °C**; Activation time: **81–90 min**	characterized by moderate pyrolysis conditioning and short, high-temperature activation



(a)

(b)

(c)

Fig. 5. Robust Pareto screening in the activation–temperature design space for each agent. (a) Agent: KOH. (b) Agent: $(NH_3)_2PO_4$. (c) Agent: $ZnCl_2$.

In the two-step route, the predicted optimal windows more clearly reflect a precarbonization priming effect, whereby the pyrolysis stage reshapes the feasible downstream activation regime. Figure 6 shows the corresponding Pareto-preferred regions in the activation-temperature design space, and the condensed route-specific operating windows are listed in Table 4. Compared with one-step activation, the two-step process introduces an additional layer of process control through pyrolysis temperature and duration, so a compact main-text summary is particularly useful for linking the graphical Pareto patterns to practically interpretable operating conditions. The full variable bounds, including feedstock-related constraints and heating-rate ranges, are provided in Table S3.

For $(NH_3)_2PO_4$4, the two-step optimum is associated with moderate impregnation intensity, a relatively strong pyrolysis stage, and subsequent activation at high temperature but controlled duration. This combination suggests that the carbon skeleton is first effectively preconditioned during pyrolysis and then further opened during activation without relying on a prolonged residence time. The result is a coherent process window that supports simultaneous gains in the SBET and VT while moderating the Yield penalty, consistent with the more stable Pareto structure observed in Figure 5.

For KOH, the two-step route differs noticeably from the one-step case. The preferred window combines an elevated soaking temperature with an extended pyrolysis stage, followed by a shorter activation step at intermediate temperature. This shift indicates that stronger precarbonization partly replaces the need for prolonged downstream etching, yielding a more controlled pathway toward pore development. In practical terms, the KOH optimum under two-step processing is better interpreted as a “preform first, activate briefly” strategy rather than the longer one-step etching pattern.

For $ZnCl_2$, the Pareto-preferred region again reflects a stable VT-oriented compromise, but here, the role of pyrolysis becomes more explicit. The recommended window combines relatively modest impregnation with mid-to-high pyrolysis conditioning and a short, high-temperature activation stage. This implies that once a suitable mesostructural precursor state is established during pyrolysis, the subsequent activation can be kept short while still preserving favorable pore-volume development. Taken together, the two-step results reinforce that the optimal process window is jointly controlled by the activating agent and route and that pyrolysis acts as a key upstream lever for shaping the final activation requirements.

These conclusions are consistent with the calibration and parity evidence. They indicate that the modeling framework not only achieves robust predictive performance but also offers coherent, implementable guidance on the kinetic trade-offs among activating agent, activation severity, pore structure, and yield.

4 Conclusions

The objective of this work was to develop a robust and interpretable modeling framework for predicting key activated-carbon properties (SBET, VT, and Yield) from tabular, small-sample, and heterogeneous process data while enabling route-/agent-specific process-window screening and multiobjective optimization under practical trade-offs. This objective was achieved by proposing AC-ResKD (Activated-Carbon Residual Knowledge Distillation), a mechanism-enhanced multitask distillation model built on a shared ResDNN backbone and dual-channel priors from prediction-level distillation (PD-KD) and teacher-aware distillation (TA-KD). By embedding key process factors (activating agent, temperature/time, impregnation, heating rate, and precursor composition) together with high-level teacher predictions, AC-ResKD establishes an interpretable input–output mapping and supports calibrated multihead learning for SBET, VT, and Yield.

The results indicate that knowledge distillation significantly enhances target-dependent performance under repeated-split evaluation (mean and 95% confidence intervals over 20 random 80/20 splits). Route-separated training consistently exhibits higher monotonicity and better calibration than combined training, and the two-step route shows the clearest precarbonization “priming” signatures in VT (higher R^2 and lower RMSE), consistent with pyrolysis presetting the carbon skeleton and pore-size spectrum. For SBET and VT, AC-ResKD achieves competitive—often higher—accuracy than the literature reference (Zou et al., 2024) when evaluated in original units, while Yield remains the most challenging target, motivating route-specific modeling and uncertainty-aware screening for actionable recommendations.

Interpretability analyses provide convergent evidence that teacher outputs, activating agent, and thermal parameters jointly dominate the input–output mapping for SBET, VT, and Yield. In one-step activation, the impregnation stage emerges as the primary determinant of accessible pore volume; in two-step activation, pyrolysis variables rise markedly in importance, aligning with mechanistic expectations. Calibration and parity plots corroborate these rankings, indicating higher linearity and lower bias for the two-step route and target-specific advantages of TA-KD and PD-KD.

Moreover, robust Pareto screening combined with quantile-window extraction identifies agent- and route-specific optimal operating ranges under multiobjective trade-offs. These intervals were subsequently adjusted to executable bounds, delineating coherent temperature–dose–time subdomains within which SBET/VT can be jointly improved while controlling yield penalties, thereby providing actionable guidance for scale-up.

The proposed approach satisfies the request for robust prediction of activated-carbon properties in complex process spaces, supports coordinated optimization of specific surface area and pore volume under yield constraints, enables screening of optimal process windows and route selection across activating agents, and can be further leveraged for parameter tuning and inverse design in activated-carbon manufacturing.

Conflict of interest

San ZHANG, Sisi LI, and Zhiwu WANG declare that they have no conflict of interest.

Data availability

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

Reference

Belluati M, Tabasso S, Calcio Gaudino E, et al., 2024. Biomass-derived carbon-based catalysts for lignocellulosic biomass and waste valorisation: a circular approach. Green Chemistry, 26(15):8642–8668. https://doi.org/10.1039/D4GC00606B

Caruana R, 1997. Multitask learning. Machine learning, 28(1):41–75.

Chairunnisa, Yu H, Saren S, et al., 2024. A review of recent advances in sustainable preparation of high-performing activated carbon for dehumidification technology. Journal of Materials Science, 59(43):20121–20156. https://doi.org/10.1007/s10853-024-10265-8

Devi R, Kumar V, Kumar S, et al., 2023. Recent advancement in biomass-derived activated carbon for waste water treatment, energy storage, and gas purification: a review. Journal of Materials Science, 58(30):12119–12142.

Evans MJB, Halliop E, MacDonald JAF, 1999. The production of chemically-activated carbon. Carbon, 37(2):269–274.

Fu J, Kang Q, Ao W, et al., 2023. Comparison and analysis of one-and two-step activation for preparation of activated carbon from furfural residues. Biomass conversion and biorefinery, 13(6):4681–4694.

Giudicianni P, Gargiulo V, Grottola CM, et al., 2021. Inherent Metal Elements in Biomass Pyrolysis: A Review. Energy & Fuels, 35(7):5407–5478. https://doi.org/10.1021/acs.energyfuels.0c04046

Gorishniy Y, Rubachev I, Khrulkov V, et al., 2021. Revisiting deep learning models for tabular data. Advances in neural information processing systems, 34:18932–18943.

Heidarinejad Z, Dehghani MH, Heidari M, et al., 2020. Methods for preparation and activation of activated carbon: a review. Environmental Chemistry Letters, 18(2):393–415. https://doi.org/10.1007/s10311-019-00955-0

Ibrahim AF, Hussein MA, 2025. Leveraging machine learning for prediction and optimization of texture properties of sustainable activated carbon derived from waste materials. Scientific Reports, 15(1):11313. https://doi.org/10.1038/s41598-025-95061-3

Kadra A, Lindauer M, Hutter F, et al., 2021. Well-tuned simple nets excel on tabular datasets. Advances in neural information processing systems, 34:23928–23941.

Kaur M, Mittal N, Charak A, et al., 2023. Rice husk derived activated carbon for the adsorption of scarlet rr an anionic disperse dye.

Kolbadinejad S, Mashhadimoslem H, Ghaemi A, et al., 2022. Deep learning analysis of Ar, Xe, Kr, and O2 adsorption on Activated Carbon and Zeolites using ANN approach. Chemical Engineering and Processing - Process Intensification, 170:108662. https://doi.org/10.1016/j.cep.2021.108662

Kundu S, Khandaker T, Anik M.A.-A.M., et al., 2024. A comprehensive review of enhanced CO_2 capture using activated carbon derived from biomass feedstock. RSC Advances, 14(40):29693–29736. https://doi.org/10.1039/D4RA04537H

Lamb WF, Grubb M, Diluiso F, et al., 2022. Countries with sustained greenhouse gas emissions reductions: an analysis of trends and progress by sector. Climate Policy, 22(1):1–17.

Li G, Xue B, Yu H, et al., 2025a. Performance improvement of waste heat upgrading adsorption heat pump by employing copper oxide-loaded composite zeolites for high-temperature steam generation. Colloids and Surfaces A: Physicochemical and Engineering Aspects, 711:136330. https://doi.org/10.1016/j.colsurfa.2025.136330

Li G, Yu H, Ji D, et al., 2025b. Pine cone-based activated carbon via dual physical activation for efficient carbon dioxide capture: Experimental and molecular simulation studies. Energy, 328:136506. https://doi.org/10.1016/j.energy.2025.136506

Li X, Huang Z, Shao S, et al., 2024. Machine learning prediction of physical properties and nitrogen content of porous carbon from agricultural wastes: Effects of activation and doping process. Fuel, 356:129623.

Liao M, Kelley SS, Yao Y, 2019. Artificial neural network based modeling for the prediction of yield and surface area of activated carbon from biomass. Biofuels, Bioproducts and Biorefining, 13(4):1015–1027. https://doi.org/10.1002/bbb.1991

Lu Y, Zhang S, Yin J, et al., 2017. Mesoporous activated carbon materials with ultrahigh mesopore volume and effective specific surface area for high performance supercapacitors. Carbon, 124:64–71.

Merchant A, Batzner S, Schoenholz SS, et al., 2023. Scaling deep learning for materials discovery. Nature, 624(7990):80–85. https://doi.org/10.1038/s41586-023-06735-9

Muhammed F, Lavaggi T, Advani S, et al., 2021. Influence of material and process parameters on microstructure evolution during the fabrication of carbon–carbon composites: a review. Journal of Materials Science, 56(32):17877–17914.

Othman NB, Kim T, Imamura A, et al., 2013. Investigation on H2S Removal Factors of Activated Carbons Derived from Waste Palm Trunk.

Przybek A, Setlak K, Piątkowski J, et al., 2025. Alkaline-Activated Materials for CO_2 Capture–Literature Review, Own observations, and Future Perspectives.

Reiser P, Neubert M, Eberhard A, et al., 2022. Graph neural networks for materials science and chemistry. Communications Materials, 3(1):93.

Ruder S, 2017. An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098.

Sevilla M, Mokaya R, 2014. Energy storage applications of activated carbons: supercapacitors and hydrogen storage. Energy Environ. Sci., 7(4):1250–1280. https://doi.org/10.1039/C3EE43525C

Shukla AK, Alam J, Mallik S, et al., 2024. Optimization and prediction of dye adsorption utilising cross-linked chitosan-activated charcoal: response surface methodology and machine learning. Journal of Molecular Liquids, 411:125745.

Song Y, Li J, Chi D, et al., 2025. AI-driven advances in metal–organic frameworks: from data to design and applications. Chemical Communications, 61(82):15972–16001.

Sosa JA, Laines JR, García DS, et al., 2023. Activated carbon: A review of residual precursors, synthesis processes, characterization techniques, and applications in the improvement of biogas. Environmental Engineering Research, 28(3):220100.

Yu H, Mikšík F, Thu K, et al., 2024. Characterization and optimization of pore structure and water adsorption capacity in pinecone-derived activated carbon by steam activation. Powder Technology, 431:119084. https://doi.org/10.1016/j.powtec.2023.119084

Yu H, Seo SW, Mikšík F, et al., 2023. Effects of temperature and humidity ratio on the performance of desiccant dehumidification system under low-temperature regeneration. Journal of Thermal Analysis and Calorimetry, 148(8):3045–3058. https://doi.org/10.1007/s10973-022-11368-7

Yuan X, Suvarna M, Low S, et al., 2021. Applied machine learning for prediction of CO2 adsorption on biomass waste-derived porous carbons. Environmental Science & Technology, 55(17):11925–11936.

Yudha CS, Hutama AP, Gustiana, H.S.E.A., et al., 2024. An Efficient Nypa Fiber Waste Conversion to Activated Carbon for Li-ion Battery Anode Material.

Zhang K, Zhong S, Zhang H, 2020. Predicting aqueous adsorption of organic compounds onto biochars, carbon nanotubes, granular activated carbons, and resins with machine learning. Environmental Science & Technology, 54(11):7008–7018.

Zhao P, Zhou J, Li G, et al., 2025. Prediction of physisorption on zeolites using Graph Integrated Adsorption Network and spline data augmentation. Thermal Science and Engineering Progress, 66:104047. https://doi.org/10.1016/j.tsep.2025.104047

Zhou Z, Yu H, Li G, et al., 2026. Material–scenario coupled techno-economics comparison of carbon capture adsorbents: Feasibility and suitability across various contexts. Energy Conversion and Management, 348:120589. https://doi.org/10.1016/j.enconman.2025.120589

Zou R, Yang Z, Zhang J, et al., 2024. Machine learning application for predicting key properties of activated carbon produced from lignocellulosic biomass waste with chemical activation. Bioresource Technology, 399:130624. https://doi.org/10.1016/j.biortech.2024.130624

Electronic supplementary materials

Section S1

中文概要

题　目：基于机制增强的多任务蒸馏技术用于生物质活性炭的预测性、可解释性设计

作　者： Peng ZHAO[1], Jiahui ZHOU[2], Guangyao LI[1,3], Hao YU[1,3], Dongxu JI[3], Takahiko MIYAZAKI[1,4], and Kyaw THU[1,4]

机　构： [1]Department of Advanced Environmental Science and Engineering, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, Fukuoka 816-8580, Japan; [2]Department of Information Science and Technology, Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan; [3]School of Science and Engineering, Chinese University of Hong Kong, Shenzhen 518172, China; [4]Research Center for Next Generation Refrigerant Properties (NEXT-RP), International Institute for Carbon-Neutral Energy Research (I2CNER), Kyushu University, Fukuoka, Japan

目　的： 针对生物质基活性炭设计中存在的合成路线异质性强以及比表面积（SBET）、总孔容（VT）和收率（Yield）之间多目标权衡显著的问题，本文旨在构建一种兼具高预测精度、可解释性和工艺决策支持能力的人工智能方法，实现对不同活化剂及一步法/两步法工艺条件下活性炭关键性能的联合预测与优化指导。

创新点： 1.提出了一种机制增强多任务蒸馏框架 AC-ResKD，以共享残差深度网络（ResDNN）为骨架，同时引入**预测层知识蒸馏（PD-KD）和教师感知知识蒸馏（TA-KD）**两类先验信息；2.面向活性炭合成中一步法与两步法机理混合、小样本、强非线性的数据特点，将工艺变量与教师模型输出联合建模，提升了多目标预测的稳定性与可解释性；3.在预测之外，进一步结合稳健 Pareto 筛选与分位数窗口提取，给出了面向不同活化剂和工艺路线的可执行操作区间，实现了从“性能预测”到“工艺决策”的延伸；4.揭示了不同路线下变量作用机制的差异：一步法中浸渍条件对 VT 更关键，两步法中热解相关变量的重要性明显上升，体现出较清晰的机理一致性。

方　法： 1.以文献数据为基础，构建了合并数据集、一步法数据集和两步法数据集三种数据配置，输入变量包括活化剂类型、活化温度/时间、浸渍条件、升温速率以及前驱体组成等工艺因素，输出为 SBET、VT 和 Yield 三个目标；2.建立共享骨干的多任务 ResDNN 学生模型，并分别与普通监督学习（Vanilla-SL）、预测层蒸馏（PD-KD）和教师感知蒸馏（TA-KD）进行对比；3.采用 20 次重复随机 80/20 划分进行训练与测试，报告原始物理单位下的平均值和 95% 置信区间，以保证结果的稳健性和可重复性；4.通过 PFI/SHAP 可解释性分析识别关键驱动因素，并结合 Pareto 优化与分位窗口提取给出多目标约束下的推荐工艺区间。

结　论： 1.所提出的 AC-ResKD 框架能够在异质、小样本活性炭合成数据上实现准确且稳健的多目标

预测，其中 TA-KD 整体表现最佳，尤其对 Yield 的提升最为显著；2.与合并建模相比，分路线建模具有更好的单调性、更低的偏差和更高的校准性，说明一步法与两步法应尽量分开建模；3.两步法在 VT 预测上表现更优，表明预碳化过程对孔结构形成具有“预设/预整形”作用；一步法中 VT 主要受浸渍阶段控制，而两步法中热解变量作用更突出；4.该方法不仅能够提供预测结果，还能给出面向不同活化剂和工艺路线的温度–剂量–时间操作窗口，为活性炭的智能设计、参数调优和逆向工艺筛选提供了可操作依据。

关键词： 活性炭合成；孔结构预测；知识蒸馏；多任务深度学习；工艺优化