Electronic supplementary materials

For 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, The Chinese University of Hong Kong, Shenzhen, Shenzhen 518172, China
[4]Research Center for Next Generation Refrigerant Properties (NEXT-RP), International Institute for Carbon-Neutral Energy Research (I2CNER), Kyushu University, Fukuoka 819-0395, Japan

Section S1 Teacher-model details

Random Forest Regressor (RF).

A random forest averages the predictions of Mregression trees $\{t_{m,k}\}_{m=1}^{M}$trained on bootstrap resamples of the training data with random feature sub-sampling. For target y_k, the RF prediction is:

$$T_k^{\mathrm{RF}}(\mathbf{x}) \;=\; \frac{1}{M}\sum_{m=1}^{M} t_{m,k}(\mathbf{x}) \tag{S1}$$

Gradient Boosting Regressor (GBRT).

Gradient boosting constructs an additive model of shallow trees $\{f_{m,k}\}_{m=1}^{M}$fitted stage-wise to the negative gradients (residuals) of a squared-error loss. For target y_k, the GBRT prediction can be written as:

$$T_k^{\mathrm{GBRT}}(\mathbf{x}) \;=\; \sum_{m=0}^{M} \nu_m\, f_{m,k}(\mathbf{x}) \tag{S2}$$

where $\nu_m$is a learning-rate (shrinkage) coefficient and $f_{0,k}$is an initial constant prediction (e.g. the mean of y_k).

Support Vector Regressor (SVR, RBF kernel).

The SVR with radial basis function (RBF) kernel represents the prediction as a weighted sum of kernel evaluations against the support vectors $\{\mathbf{x}_i\}_{i\in\mathcal{S}}$:

$$T_k^{\mathrm{SVR}}(\mathbf{x}) \;=\; \sum_{i\in\mathcal{S}}\left(\alpha_{i,k}-\alpha_{i,k}^{*}\right)K(\mathbf{x}_i,\mathbf{x})+b_k \tag{S3}$$

where $K(\mathbf{x}_i,\mathbf{x})=\exp\left(-\gamma\parallel\mathbf{x}_i-\mathbf{x}\parallel^2\right)$is the RBF kernel, and $\alpha_{i,k},\alpha_{i,k}^{*}$are dual variables determined by the ε-insensitive SVR optimisation.

k-Nearest Neighbours Regressor (kNN).

For a query point $\mathbf{x}$, the kNN regressor averages the labels of the Kclosest training samples (with respect to a chosen distance metric):

$$T_k^{\mathrm{kNN}}(\mathbf{x}) \;=\; \frac{1}{Z(\mathbf{x})}\sum_{i\in\mathcal{N}_K(\mathbf{x})}w_i(\mathbf{x})\,y_{k,i} \tag{S4}$$

where $\mathcal{N}_K(\mathbf{x})$denotes the set of Knearest neighbours in the training data, $w_i(\mathbf{x})$is either a uniform weight ($w_i\equiv 1$) or a distance-based weight, and $Z(\mathbf{x})=\sum_{i\in\mathcal{N}_K(\mathbf{x})}w_i(\mathbf{x})$ is a normalization factor.

Three-layer MLP regressor

The shallow neural teacher is a fully-connected three-layer MLP with ReLU nonlinearities. For an input $\mathbf{x}$, the hidden activations are:

$$\mathbf{h}^{(1)}=\sigma\left(W^{(1)}\mathbf{x}+\mathbf{b}^{(1)}\right),\mathbf{h}^{(2)}=\sigma\left(W^{(2)}\mathbf{h}^{(1)}+\mathbf{b}^{(2)}\right),\mathbf{h}^{(3)}=\sigma\left(W^{(3)}\mathbf{h}^{(2)}+\mathbf{b}^{(3)}\right) \tag{S5}$$

where $\sigma(\cdot)$is the ReLU activation. The scalar prediction for target $y_k$is obtained by a linear head on the last hidden layer,

$$T_k^{\mathrm{MLP}}(\mathbf{x}) \;=\; \mathbf{w}_k^{\top}\mathbf{h}^{(3)}+b_k \tag{S6}$$

All teacher models operate on the *same* preprocessed feature vector $\mathbf{x}$, and are trained separately for each target y_k.

In addition to individual models, this work also considers simple unweighted ensembles built from 2–4 heterogeneous teachers. For a given target y_k, let $\mathcal{C}_k=\{c_1,\ldots,c_R\}$denote a subset of candidate base models (e.g. RF+GBRT+SVR, or SVR+MLP), where $R\in\{2,3,4\}$. The corresponding equal-weight ensemble prediction is:

$$T_k^{\mathrm{ens}}(\mathbf{x}) \;=\; \frac{1}{R}\sum_{r=1}^{R}T_k^{(c_r)}(\mathbf{x}) \tag{S7}$$

where $T_k^{(c_r)}(\mathbf{x})$ denotes the prediction of the r -th base model in the ensemble for target k . These ensembles are computed purely at prediction time by averaging the outputs of the constituent teachers; no additional training is required beyond fitting the individual base models.

Section S2 External-dataset benchmarking

To address concerns about robustness beyond a single curated dataset, we additionally evaluated our framework on an independent activated-carbon dataset from Ibrahim [28], which compiles texture-property targets including surface area (SA), total pore volume (TPV), and micropore volume (MPV). This dataset differs from Zou et al. in descriptor schema and curation rules, providing a complementary test of cross-source generalization.

Table S1 The model results on external dataset.

Method	Target	R^2 (mean [95%CI])	Pearson (mean [95%CI])	RMSE (mean [95%CI])
Vanilla-SL	SA (m^2/g)	0.722 [0.676, 0.768]	0.852 [0.824, 0.880]	315.486 [287.098, 343.875]
	TPV (cm^3/g)	0.735 [0.704, 0.765]	0.866 [0.847, 0.885]	0.168 [0.159, 0.177]
	MPV (cm^3/g)	0.707 [0.663, 0.751]	0.842 [0.814, 0.870]	0.121 [0.112, 0.129]
PD-KD	SA (m^2/g)	0.742 [0.700, 0.784]	0.865 [0.839, 0.891]	301.902 [276.502, 327.302]
	TPV (cm^3/g)	0.747 [0.716, 0.777]	0.874 [0.855, 0.893]	0.164 [0.155, 0.173]
	MPV (cm^3/g)	0.725 [0.683, 0.767]	0.854 [0.827, 0.880]	0.117 [0.109, 0.126]
TA-KD	SA (m^2/g)	0.759 [0.717, 0.801]	0.881 [0.855, 0.907]	289.527 [264.158, 314.896]
	TPV (cm^3/g)	0.760 [0.734, 0.785]	0.889 [0.873, 0.905]	0.161 [0.154, 0.169]
	MPV (cm^3/g)	0.748 [0.708, 0.788]	0.871 [0.847, 0.896]	0.113 [0.105, 0.121]
Teacher	SA (m^2/g)	0.771 [0.748, 0.793]	0.895 [0.882, 0.908]	280.374 [263.054, 297.694]
	TPV (cm^3/g)	0.758 [0.739, 0.776]	0.889 [0.878, 0.900]	0.161 [0.154, 0.168]
	MPV (cm^3/g)	0.755 [0.729, 0.781]	0.875 [0.860, 0.891]	0.112 [0.105, 0.119]

Table S1 summarizes the results. Across all three targets, knowledge distillation yields consistent improvements over Vanilla-SL, with TA-KD providing the strongest student performance (SA: R^2 = 0.759 [0.717, 0.801]; TPV: R^2 = 0.760 [0.734, 0.785]; MPV: R^2 = 0.748 [0.708, 0.788]). While absolute R^2 values may differ from those reported in the source paper due to differences in preprocessing and evaluation settings, the relative ranking and stability of gains are consistent, supporting that AC-ResKD is not specific to a single dataset.

Section S3 Additional calibration analysis

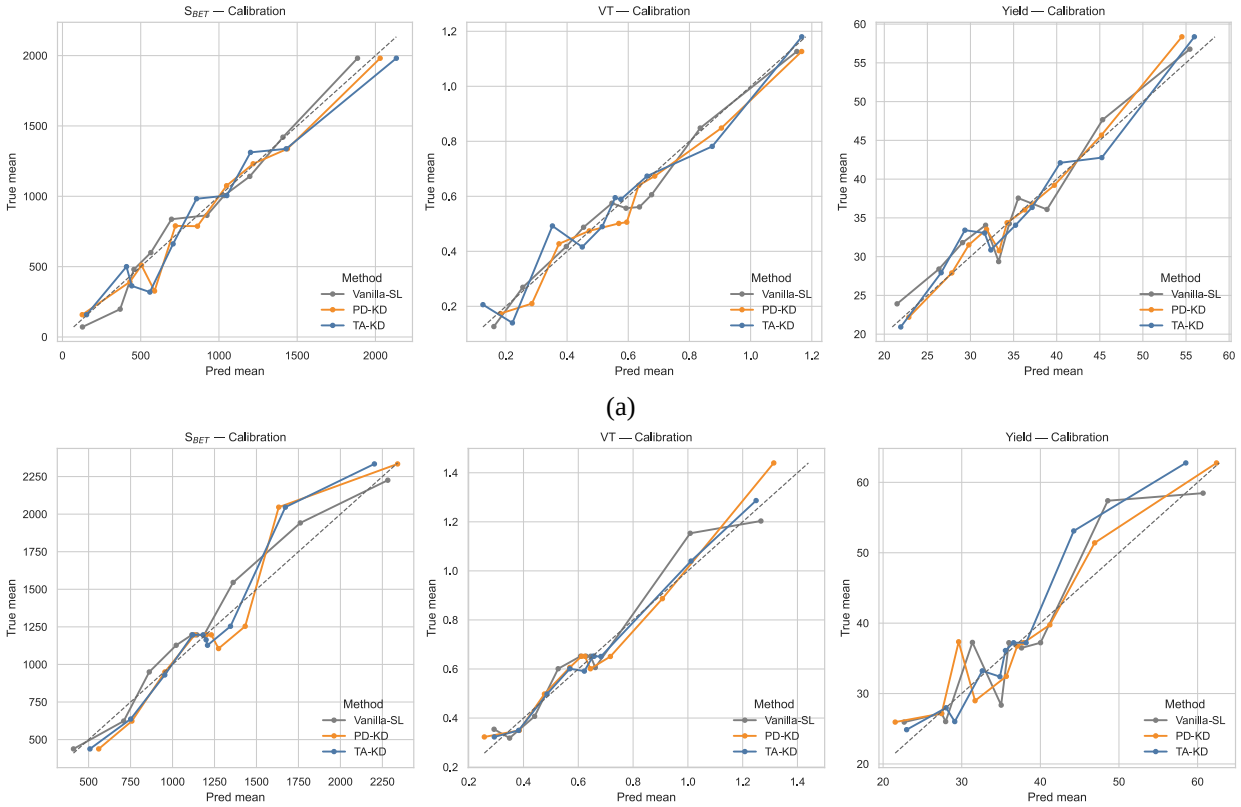


Fig.S1 Calibration curves on the one-step and two-step configurations. (a) *one-step* calibration curves for each target. (b) *two-step* calibration curves for each target

Figure S1 presents calibration curves based solely on the one-step and two-step configurations. The combined setting is not considered here due to inferior performance. Calibration curves are obtained by partitioning predictions into ten equal-frequency bins along the x-axis and plotting, for each bin, the mean prediction against the mean observation, connected by line segments. Proximity to the y = x diagonal indicates low bias and a more reliable monotonic mapping within that interval. Segments above or below the diagonal correspond to systematic overestimation or underestimation, respectively. S-shaped patterns or slope contraction indicate nonlinear mismatch or regression to the mean.

A comparison of one-step versus two-step calibration reveals superior monotonicity and linearity in the two-step setting, with slopes closer to 1 and smaller interval bias—particularly for VT and Yield. The one-step setting exhibits modest mid-range or low-end deviations, although overall bias remains limited. At the extremes, two-step is more stable in the high-value regime (S_{BET}, VT), and PD-KD largely removes upper-tail compression for VT. In one-step, S_{BET} performs best under Vanilla-SL at the upper tail, suggesting that teacher smoothing under-captures extreme microstructural variation for this target. The target-dependent preference across distillation paradigms is clear: in one-step, S_{BET} favors Vanilla-SL, VT favors TA-KD, and Yield favors PD-KD; in two-step, S_{BET} is more stable under Vanilla-SL/TA-KD, VT strongly favors PD-KD, and Yield favors TA-KD.

Section S4 One-step operating windows

To complement the condensed main-text summary, this section provides the full one-step operating-window bounds and an extended interpretation of the agent-specific Pareto-preferred regimes. The ranges reported below retain the complete variable set, including feedstock-composition descriptors and auxiliary processing variables, and therefore provide a fuller view of the feasible design space underlying the main-text recommendations.

Table S2 Optimal operating window under different activating agents by one-step activation

Input Factor	Best Interval of H_3PO_4	Best Interval of KOH	Best Interval of $ZnCl_2$
VM	[74.56, 85.69]	[59.32, 85.42]	[59.67, 85.18]
Ash	[2.85, 14.23]	[2.05, 14.00]	[5.15, 14.50]
FC	[11.44, 18.39]	[11.99, 30.52]	[12.05, 32.25]
agent/sample	[1.60, 7.95]	[0.83, 6.45]	[0.43, 5.81]
soaking time (min)	[990.86, 2830.70]	[1132.89, 2786.93]	[703.24, 2754.49]
soaking temp (°C)	[25.54, 68.53]	[26.49, 81.62]	[29.87, 98.41]
Activation temperature (°C)	[369.93, 884.24]	[357.88, 699.63]	[361.22, 884.047]
Activation time(min)	[56.95, 159.54]	[73.03, 174.25]	[58.91, 171.67]
Heating rate (°C/min)	[2, 10]	[2, 10]	[2, 10]

The detailed one-step results indicate that the Pareto-preferred operating domains are strongly agent-dependent, even when the same three target properties (SBET, VT, and Yield) are optimized jointly. Across all three agents, the preferred regions avoid the low-temperature/low-dose corner and also avoid excessively severe operating conditions that yield only limited incremental textural gains while imposing a substantial Yield penalty. This confirms that the one-step process does not exhibit a single universal optimum; instead, each activating chemistry defines its own temperature–dose–time plateau within the broader feasible region.

For H3PO4, the robust Pareto set is associated with feedstocks characterized by relatively high volatile matter, low-to-moderate ash, and comparatively low fixed carbon. This pattern is consistent with the well-known tendency of phosphoric acid activation to favor a more plasticizable precursor skeleton during dehydration and crosslinking. In the process-variable space, the one-step H3PO4 window is distinguished by comparatively high agent loading together with long soaking duration, whereas soaking temperature remains only moderate. This suggests that, for H3PO4, the quality and uniformity of chemical penetration are more strongly governed by dose and residence time than by aggressive thermal pretreatment at the impregnation stage. The downstream activation regime centers on relatively high temperature with moderate duration, which supports stable VT development while limiting collapse of the developing pore skeleton. Overall, the H3PO4 route in one-step activation can be interpreted as a “high-dose, thorough-impregnation, moderate-duration activation” regime that prioritizes balanced improvement of SBET and VT under controlled Yield penalty.

For KOH, the detailed operating window reflects the characteristic efficiency of strong-base activation. The preferred feedstock region shifts toward higher fixed carbon and moderate ash content, consistent with the expectation that potassium-driven reactions develop microporosity more effectively on a relatively rigid carbon backbone. In contrast to H3PO4, the KOH optimum is achieved at substantially lower activation temperature, although it still requires adequate impregnation and a slightly longer activation duration. This indicates that KOH can generate a high-platform microporous structure without requiring the highest thermal severity, but it also implies a narrower tolerance with respect to Yield preservation. Within the Pareto-preferred region, increases in severity beyond the recommended window are more likely to translate into over-etching and accelerated carbon loss than

into proportionally large improvements in SBET or VT. The one-step KOH regime may therefore be regarded as a more efficient but more severity-sensitive route.

For $ZnCl_2$, the detailed window shows a different balance between dosage, soaking conditions, and downstream activation. Relative to the other two agents, the $ZnCl_2$ optimum is less dose-intensive but more responsive to soaking temperature, suggesting that enhanced solute mobility and temperature-assisted site accessibility are especially relevant for this salt-based route. The preferred activation regime lies in a mid-to-high temperature window with moderate duration, and the resulting Pareto-preferred region is more consistently aligned with stable VT enhancement than with the pursuit of extreme SBET values. This behavior is compatible with the mesopore-oriented amplification typically associated with $ZnCl_2$ activation chemistry. In practical terms, the one-step $ZnCl_2$ window can be viewed as a VT-oriented, temperature-assisted impregnation route with moderate activation severity.

Taken together, the one-step detailed windows support the main-text conclusion that activating agent identity is not merely a categorical descriptor but a primary determinant of the feasible operating regime. H3PO4 favors thorough impregnation and relatively high activation temperature, KOH favors efficient lower-temperature etching with stronger Yield sensitivity, and $ZnCl_2$ favors a more VT-oriented compromise supported by elevated soaking temperature. These detailed ranges may be used when finer process specification is needed beyond the condensed summary presented in the main text.

Section S5 Two-step operating windows

This section provides the full two-step operating-window bounds and an extended interpretation of the route-specific Pareto-preferred regimes. Compared with one-step activation, the two-step process introduces additional control through pyrolysis temperature, pyrolysis duration, and pyrolysis heating conditions, making the full-variable description particularly useful for understanding how upstream skeleton preconditioning reshapes the downstream activation requirements.

Table S3 Optimal operating window under different activating agents by two-step activation

Input Factor	Best Interval of H_3PO_4	Best Interval of KOH	Best Interval of $ZnCl_2$
VM	[63.26, 82.25]	[62.81, 82.26]	[62.79, 81.51]
Ash	[0.83, 20.08]	[0.69, 19.87]	[1.32, 22.06]
FC	[10.92, 26.54]	[11.48, 26.80]	[11.46, 27.04]
Agent/sample	[0.29, 3.74]	[0.31, 3.89]	[0.25, 3.46]
Soaking time (min)	[199.38, 1410.93]	[287.80, 1405.10]	[378.43, 1412.68]
Soaking temperature (°C)	[30.62, 115.74]	[30.88, 116.15]	[29.94, 114.23]
Pyrolysis temperature (°C)	[236.98, 986.48]	[254.93, 989.01]	[297.55, 985.18]
Pyrolysis time (min)	[48.19, 524.40]	[69.27, 654.15]	[46.62, 561.18]
Pyrolysis heating rate (°C/min)	[6.37, 47.43]	[6.31, 46.07]	[6.72, 44.38]
Activation temperature (°C)	[519.19, 895.36]	[507.57, 891.91]	[528.77, 894.26]
Activation time (min)	[37.50, 117.78]	[53.68, 118.74]	[35.74, 117.53]
Activation heating rate (°C /min)	[2, 20]	[2, 20]	[2, 20]

The two-step results reveal a more explicit route-dependent structure than the one-step case, primarily because the pyrolysis stage introduces an additional degree of freedom that can precondition the evolving carbon skeleton before chemical activation. Across all three activating agents, the detailed windows are consistent with a “pre-

carbonization priming" interpretation: once the precursor structure has been sufficiently rearranged during pyrolysis, the downstream activation stage can be more tightly controlled, often with shorter duration or more focused thermal intensity than in the corresponding one-step route.

For H3PO4, the full two-step window corresponds to moderate impregnation intensity, a relatively strong pyrolysis stage, and a subsequent activation step at high temperature but controlled duration. Compared with the one-step phosphoric-acid route, the two-step process requires less extreme impregnation intensity because the precursor skeleton has already been more uniformly reorganized prior to activation. The pyrolysis window lies in a mid-to-high temperature region with moderate residence time and relatively fast heating, indicating that the role of pre-carbonization is not simply drying or stabilization, but active preparation of a dense yet etchable substrate. The subsequent activation step then operates at high temperature with limited residence time, which supports simultaneous gains in SBET and VT without requiring prolonged downstream exposure. This combination is consistent with a coordinated "preform first, open later" strategy.

For KOH, the two-step regime differs most clearly from its one-step counterpart. The preferred window combines moderate agent loading and elevated soaking temperature with a more extended pyrolysis stage, followed by an activation step at intermediate temperature and shorter duration than would typically be favored in the one-step route. This suggests that the KOH pathway benefits strongly from upstream skeleton conditioning: once the carbon framework has been sufficiently preformed, the need for long downstream etching is reduced. The practical implication is that the two-step KOH route should not be interpreted simply as "one-step KOH plus pyrolysis," but rather as a distinct strategy in which the burden of pore development is redistributed between pre-carbonization and activation. This redistribution appears to improve the controllability of SBET and VT while reducing the risk of unnecessary Yield loss associated with prolonged activation.

For $ZnCl_2$, the two-step window preserves the agent's characteristic VT-oriented behavior while making the role of pyrolysis more explicit. The detailed range indicates relatively modest impregnation intensity, a mid-to-high pyrolysis regime, and a short but high-temperature activation step. This implies that once a suitable mesostructural precursor state is established during pyrolysis, the downstream salt-assisted activation can function more as targeted amplification than as wholesale restructuring. As a result, VT enhancement remains stable within the Pareto-preferred region, while SBET gains remain moderate and Yield penalties can be kept within an acceptable operational range. In comparison with the one-step route, the two-step $ZnCl_2$ window therefore appears more structured and more dependent on upstream preparation of the precursor skeleton.

Overall, the detailed two-step windows reinforce the main-text conclusion that the final process optimum is jointly controlled by activating agent and route. In the two-step setting, pyrolysis functions as a major upstream lever that reshapes the allowable activation requirements, rather than as a neutral preprocessing step. H3PO4 favors coordinated preforming and high-temperature short activation, KOH shifts toward stronger pre-carbonization and briefer downstream etching, and $ZnCl_2$ exhibits a stable VT-oriented compromise supported by explicit pyrolysis conditioning. These detailed bounds provide a more complete reference for route-specific process design than the condensed main-text summary alone.