Supplementary Information

XPEAK: An XRD-Driven Machine Learning Platform for Predicting the Catalyst-Enhanced Dehydrogenation Peak Temperature of MgH_2

Weijie Yang [ab,*], Zhuoran Xu [ab], Jiongyang Li [c], Leping Zhang [c], Man Shu [ab], Jianghao Cai [ab], Tongao Yao [ab], Xiaotian Tang [ab], Yuxuan Liu [ab], Shiwen Luo [ab], Xuqiang Shao [c], Wenjing Wang [d], Zhengyang Gao [ab]

[a] Department of Power Engineering, School of Energy, Power and Mechanical Engineering, North China Electric Power University, Baoding, 071003, China

[b] Hebei Key Laboratory of Energy Storage and Integrated Energy Systems, North China Electric Power University, Baoding 071003, Hebei, China

[c] Department of Computer Science, North China Electric Power University, Baoding 071003, Hebei, China

[d] College of Life Sciences, Institute of Life Sciences and Green Development, Hebei University, Baoding 071002, China

***Corresponding Author:** Weijie Yang(yangwj@ncepu.edu.cn)

Contents

Supplementary Method

Method S1

Scientific literature represents an important and reliable source of data in the field of materials science. With years of accumulation, a substantial volume of experimental data is now available for constructing robust datasets. In this study, the data used for regression model development were obtained through a systematic review of 810 publications related to Mg-based solid-state hydrogen storage materials published between 2005 and 2024. The literature was retrieved from databases such as Web of Science and Google Scholar. During the initial screening, studies involving complex hydride systems (e.g., MgH_2–$LiBH_4$[1,2], MgH_2–AlH_3[3–5]), Mg-based alloys, and $Mg(BH_4)_2$[6] were excluded. A total of 453 studies focusing specifically on the catalytic regulation of hydrogen storage/release performance of MgH_2 were retained. After further excluding papers lacking differential scanning calorimetry (DSC) characterization data, 420 publications were ultimately selected as valid sources for model training.

A total of 2,011 data entries were compiled for "MgH_2 + catalyst" composite systems, each containing DSC measurements of the *Tp*. Among them, 1,349 samples additionally included the *ΔTp* before and after catalyst addition, all of which were used for regression modeling. Given the significant influence of heating rates on *Tp*[7,8], the dataset also records the heating rates for each sample, and all *ΔTp* values were calculated under identical heating conditions to ensure data comparability and physical consistency. Furthermore, to account for the effect of catalyst dosage on performance, the mass fraction of the added catalyst was also recorded for each sample[9,10].

In addition, to further advance the field of solid-state hydrogen storage, we have developed a dedicated database platform named Digital Hydrogen-S in a separate project. This platform provides processing information, performance characterization, and morphological characterization data for MgH_2 and a broader range of materials. The data from the present study has been integrated into this website, which is accessible at https://s-hydrogendataplatform.nas.cpolar.cn.

Method S2

There is a discrepancy in the sampling points between the original XRD spectra from the literature and databases, which is not conducive to subsequent modeling. To address this, the following standardization process was adopted: all spectra were unified to an equidistant grid in the 2θ range of 5° to 90° with a step size of $\Delta 2\theta = 0.02$ using linear interpolation. The interpolation formula is as follows:

$$2\theta = 2\theta_0 + \frac{(2\theta_1 - 2\theta_0)}{(I_1 - I_0)}(I - I_0) \quad (1)$$

All relative intensities were normalized to the [0, 1] range using MinMaxScaler:

$$I_{\text{scaled}} = \frac{I - \min(I)}{\max(I) - \min(I)} \quad (2)$$

In equations (1) and (2), 2θ represents the diffraction angle, and I represents the relative intensity of the spectrum.

To balance overall accuracy and computational cost, this study uniformly defines the diffraction angle range of all XRD feature matrices as 5°-90°. Linear interpolation with a step size of 0.02° is applied to ICDD XRD data, while quadratic interpolation with a step size of 0.01° is used for EXP XRD data.

To ensure good interpolation results and enhance the robustness of the program, the following design was implemented: the coordinates of the top 20 peaks with the highest relative intensities in the original spectrum were dynamically saved, and for repeated angles, the average intensity was taken to avoid interpolation anomalies. Interpolation was performed using the linear method, and the diffraction angle boundaries were extended to reduce edge effects in the interpolation. After interpolation, the intensity of the nearest neighboring grid points was replaced to preserve key structural features. For samples with missing intensity values in the standard interpolation grid, zero-padding was applied to ICDD XRD data, and boundary extension was used for EXP XRD data. Since the spectra contain all representative diffraction peaks and missing data occur mainly in non-characteristic regions, these approaches do not compromise phase representation or regression model performance.

Method S3

To elucidate the complex nonlinear relationships between material features and dehydrogenation performance, this study employs the SHapley Additive exPlanations (SHAP) method. SHAP quantifies the marginal contribution of each feature to the prediction of an individual sample, under the following axiomatic framework:

$$\phi_i(\mathbf{f}, x) = \sum_{S \subseteq \mathbf{N} \setminus \{i\}} \frac{|S|!(\mathbf{M} - |S| - 1)!}{\mathbf{M}!} \left[\mathbf{f}_{\mathbf{x}}(S \cup \{i\}) - \mathbf{f}_{\mathbf{x}}(S) \right] \tag{3}$$

In Equation (3), ϕ_i is the SHAP value of feature i, $\mathbf{M}$ is the total number of features, $\mathbf{N}$ denotes the complete set of features, and $S \subseteq N \setminus \{i\}$ represents any subset that excludes feature i.

t-SNE is a nonlinear dimensionality reduction algorithm commonly used for the visualization of high-dimensional data. The method optimizes the low-dimensional embedding by minimizing the Kullback-Leibler divergence between the two distributions. The Kullback-Leibler divergence is calculated as follows:

In Equation (4), $\boldsymbol{p_{ij}}$ represents the true distribution, and $\boldsymbol{q_{ij}}$ represents the approximate distribution

$$\mathrm{KL}(\mathbf{P} \| \mathbf{Q}) = \sum_{i \neq j} \boldsymbol{p_{ij}} \log \frac{\boldsymbol{p_{ij}}}{\boldsymbol{q_{ij}}} \tag{4}$$

Method S4

In experimental studies, XRD patterns of synthesized samples often deviate from ideal diffraction profiles due to factors such as grain size, lattice defects or stress, preferred crystallographic orientation, and instrumental noise. These deviations typically manifest as peak broadening, peak splitting, and background noise. To generate training and input data that better resemble experimental data reported in the literature, we propose a simulation method for generating pseudo-experimental XRD patterns based on physical principles and standard diffraction profiles. This method incorporates multiple factors, including microstructural characteristics, instrumental parameters, and experimental noise. A comprehensive model was developed to simulate the effects of crystallite size distribution, microstrain, thermal diffuse scattering, and instrumental broadening. The crystallite size was randomly sampled within a defined range to enhance the robustness and generalizability of the generated data.

The peak profiles were simulated using the Voigt function, which is the convolution of a Gaussian and a Lorentzian function. The mathematical expression of the Voigt function is given by:

$$V(x) = \int_{-\infty}^{\infty} G\left(x'\right) L\left(x - x'\right) dx' \tag{5}$$

The Gaussian component represents the peak broadening caused by the instrument, and its full width at half maximum (FWHM) is calculated using the Caglioti equation:

$$H_{inst} = \sqrt{U\,tan^2\,\theta + V tan\theta + W} \tag{6}$$

The Lorentzian component accounts for peak broadening caused by crystallite size and lattice strain, and is calculated as follows:

$$H_{size} = \frac{K\lambda}{D cos\,\theta} \tag{7}$$

$$H_{strain} = 4\varepsilon\, tan\,\theta \tag{8}$$

The crystallite size is generated based on a skew normal distribution, whose probability density function is given by:

$$f\left(x \mid \mu, \sigma, \alpha\right) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\alpha\left(\frac{x-\mu}{\sigma}\right)\right) \tag{9}$$

The thermal diffuse scattering effect is introduced by incorporating the Debye–Waller factor, which accounts for the intensity attenuation caused by atomic thermal vibrations:

$$I_{TDS} = I_0 \, exp\left(-B \frac{sin^2 \theta}{\lambda^2} \right) \tag{10}$$

The preferred orientation of the crystal is simulated using the following equation:

$$I_{hkl} = I_0 \cdot \frac{\beta}{\sqrt{1 + \left(\beta^2 - 1 \right) cos^2 \psi}} \tag{11}$$

A probabilistic triggering mechanism was employed to simulate peak splitting, where peak splitting occurs if the splitting probability P>0.3P > 0.3P>0.3. Background noise was diversified by incorporating Gaussian white noise, impulsive noise, and baseline drift.

Supplementary Figures

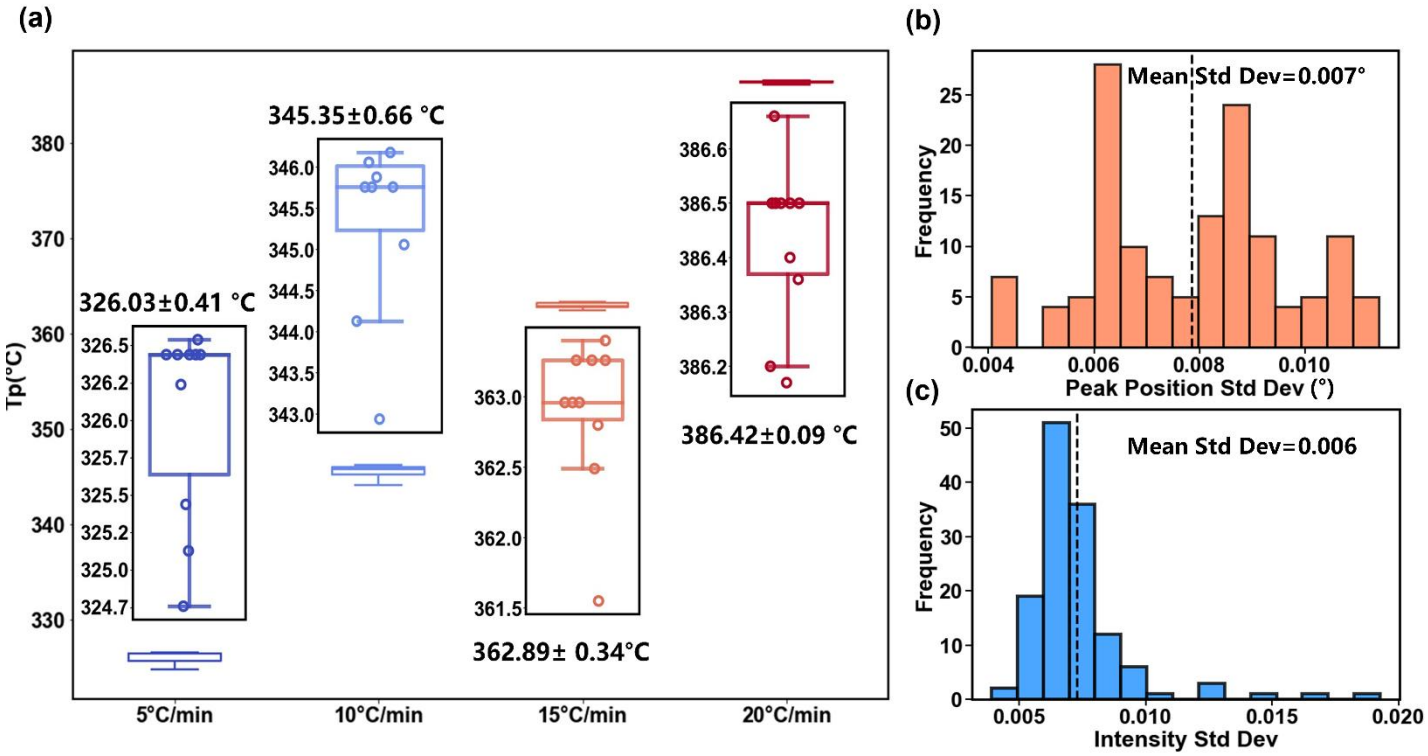

Fig. S1 (a) Analysis of the uncertainties associated with manually extracting Tp values using WebPlotDigitizer. (b) Standard deviation distribution of 2θ peak positions extracted from XRD data. (c) Standard deviation distribution of normalized peak intensity extracted from XRD data.

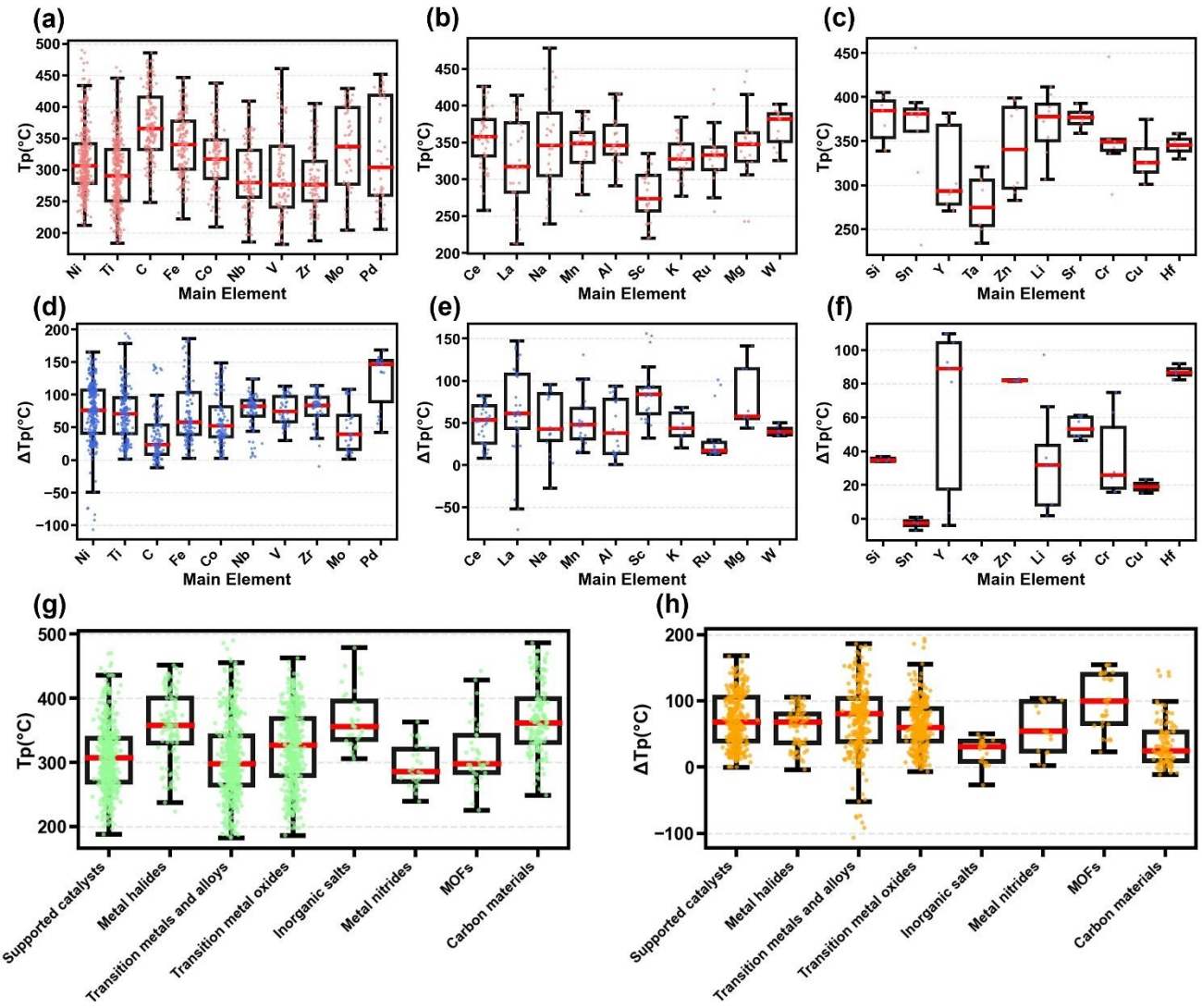

Fig. S2 (a–c) Distribution of Tp for catalysts composed of different elemental species; (d–f) distribution of ΔTp for catalysts composed of different elemental species; (g) distribution of Tp across catalyst categories; (h) distribution of ΔTp across catalyst categories.

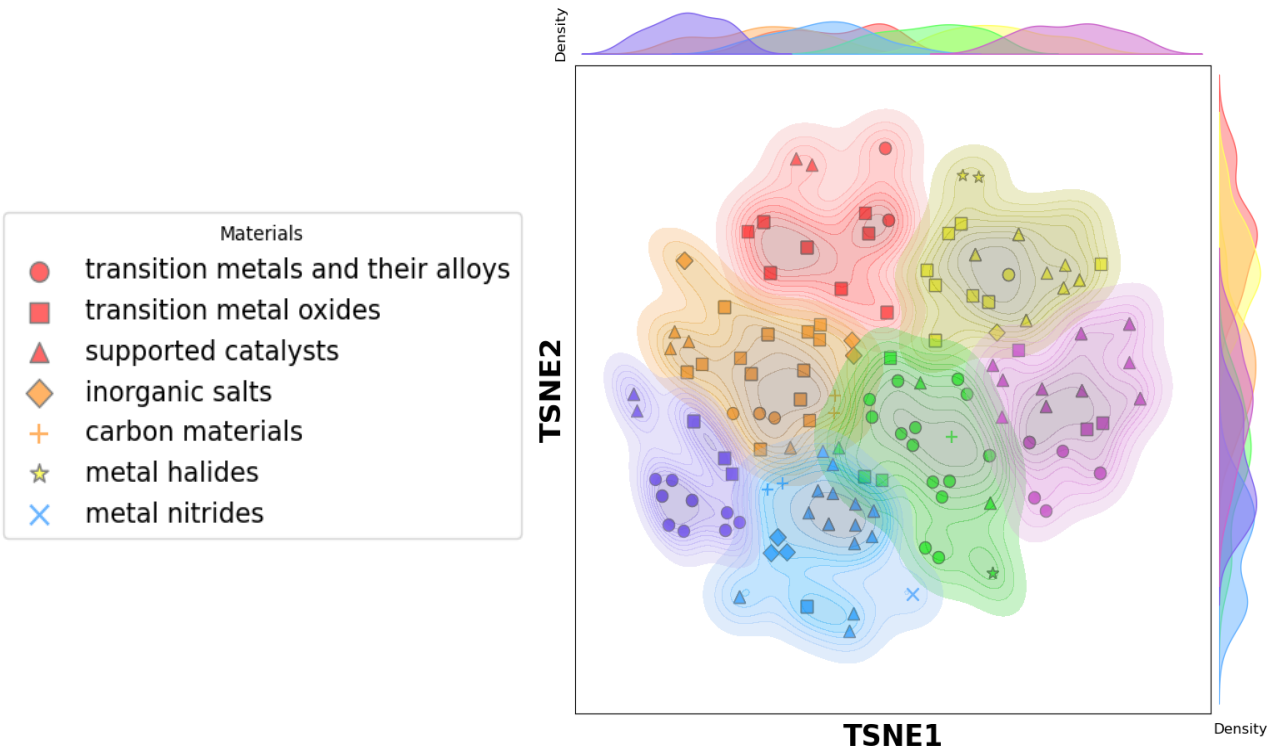

Fig. S3 Feature clustering analysis of selected samples using t-SNE and K-Means.

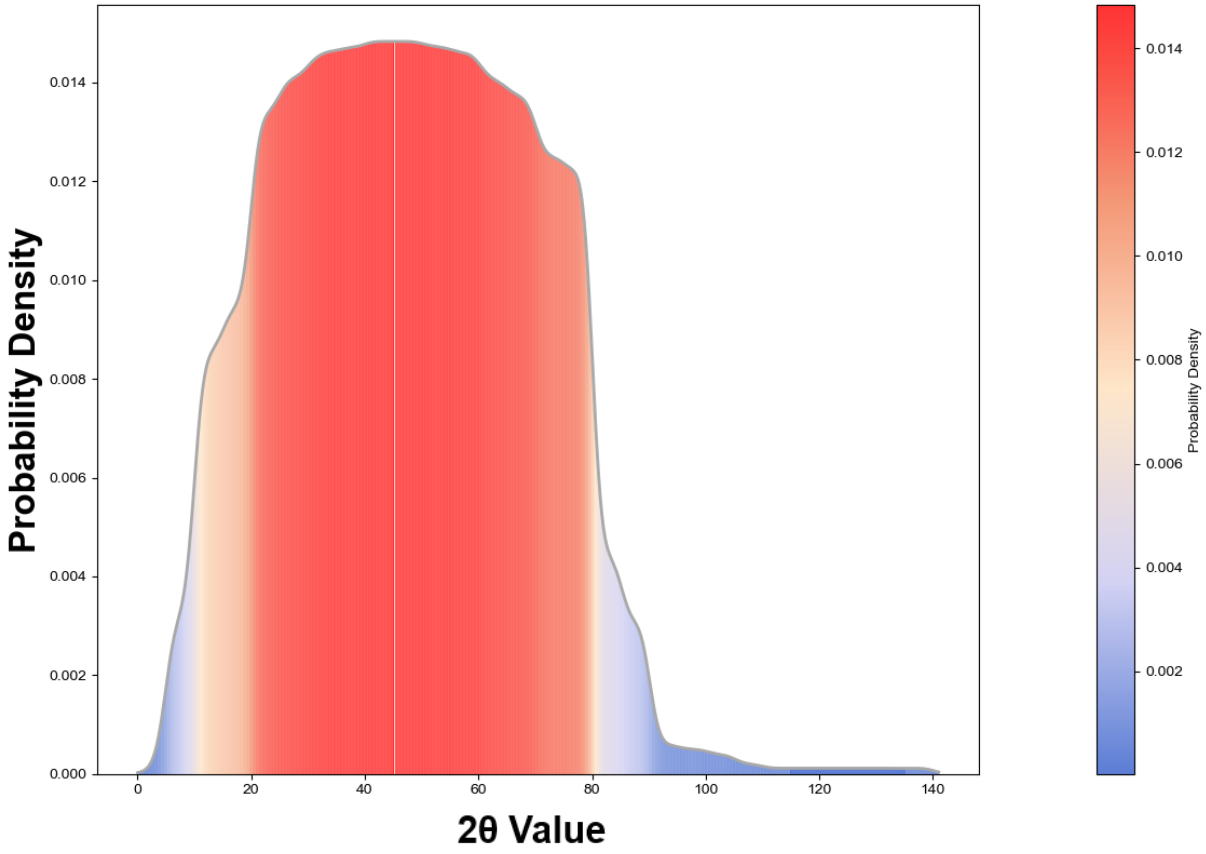

Fig. S4 Probability density distribution of diffraction angles corresponding to relative diffraction intensities in the catalyst XRD data from the database.

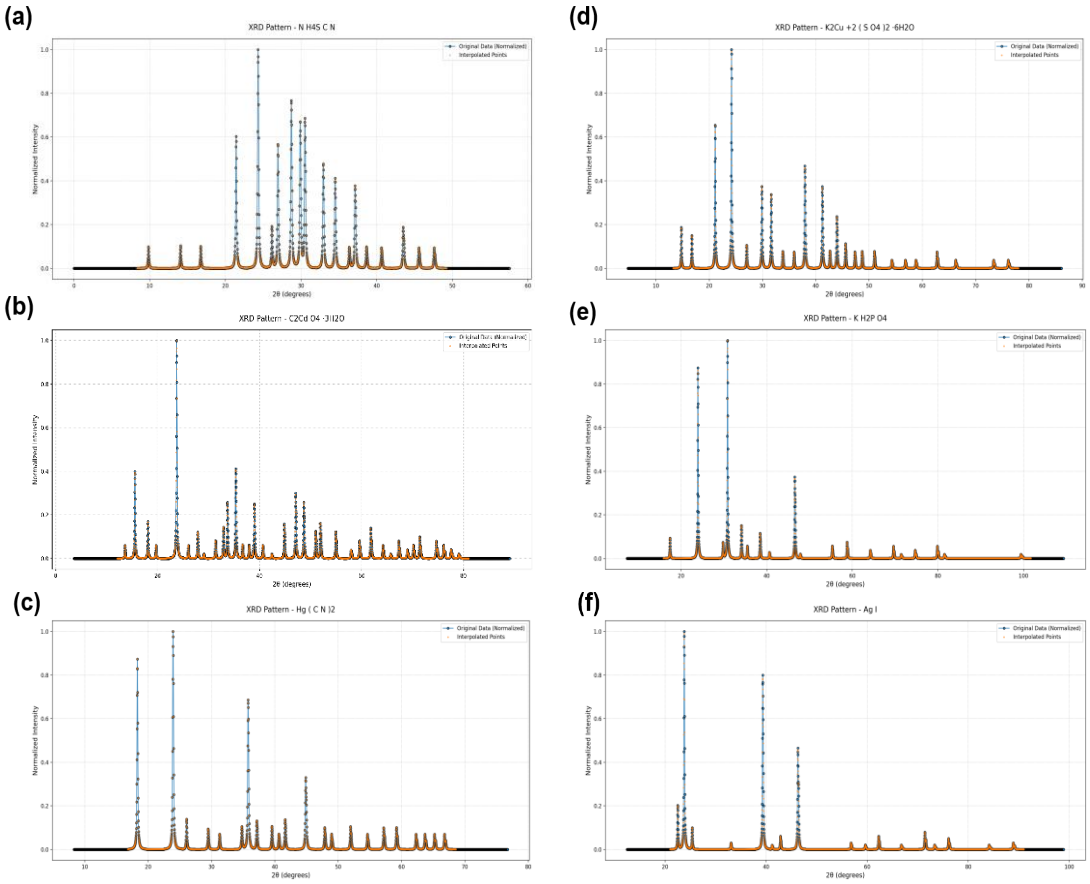

Fig. S5 Interpolation performance of randomly selected samples from the prediction set. The figure compares the original XRD patterns with those reconstructed by linear interpolation after normalization.

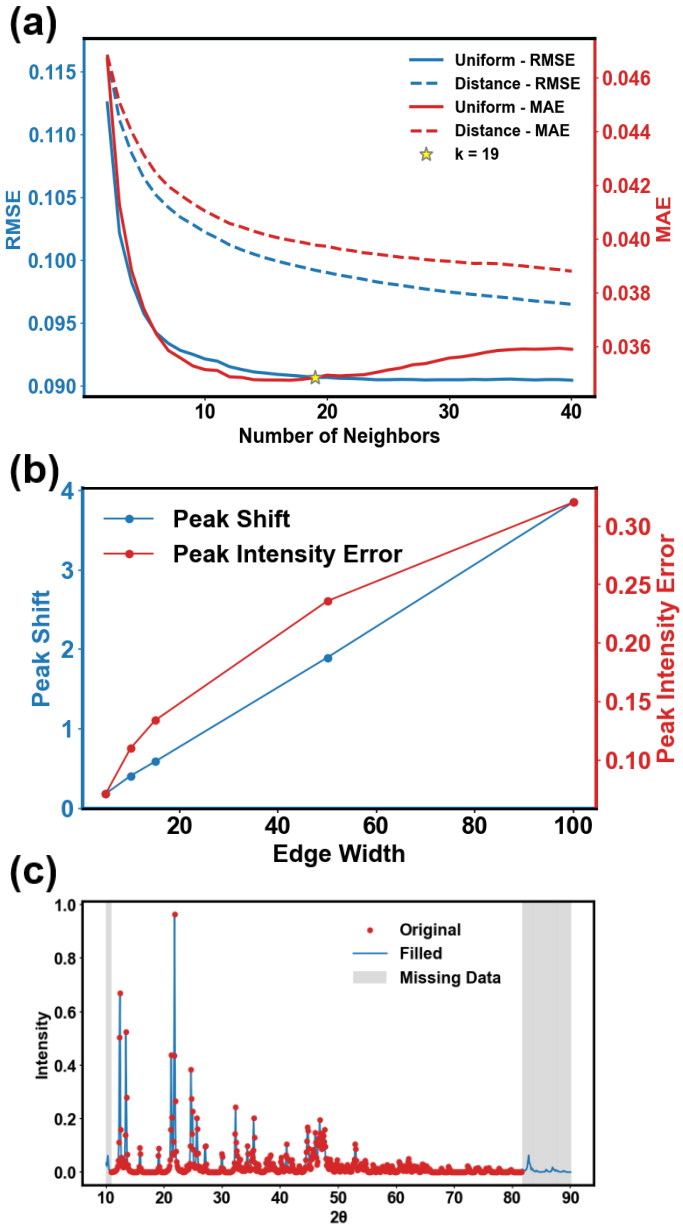

Fig. S6 Performance of the KNN imputation model combined with a weighted loss function for fitting missing boundary points. (a) MAE (red) and RMSE (blue) trends with varying neighbor counts using Uniform (solid lines) and Distance (dashed lines) sample distance metrics. (b) Trends of peak position shifts and peak intensity deviations as boundary missing intensity varies. (c) Imputation example for a random sample: red points represent original data, blue solid line is the fitted spectrum after boundary supplementation, and the gray shaded area indicates the region of missing data in the original spectrum.

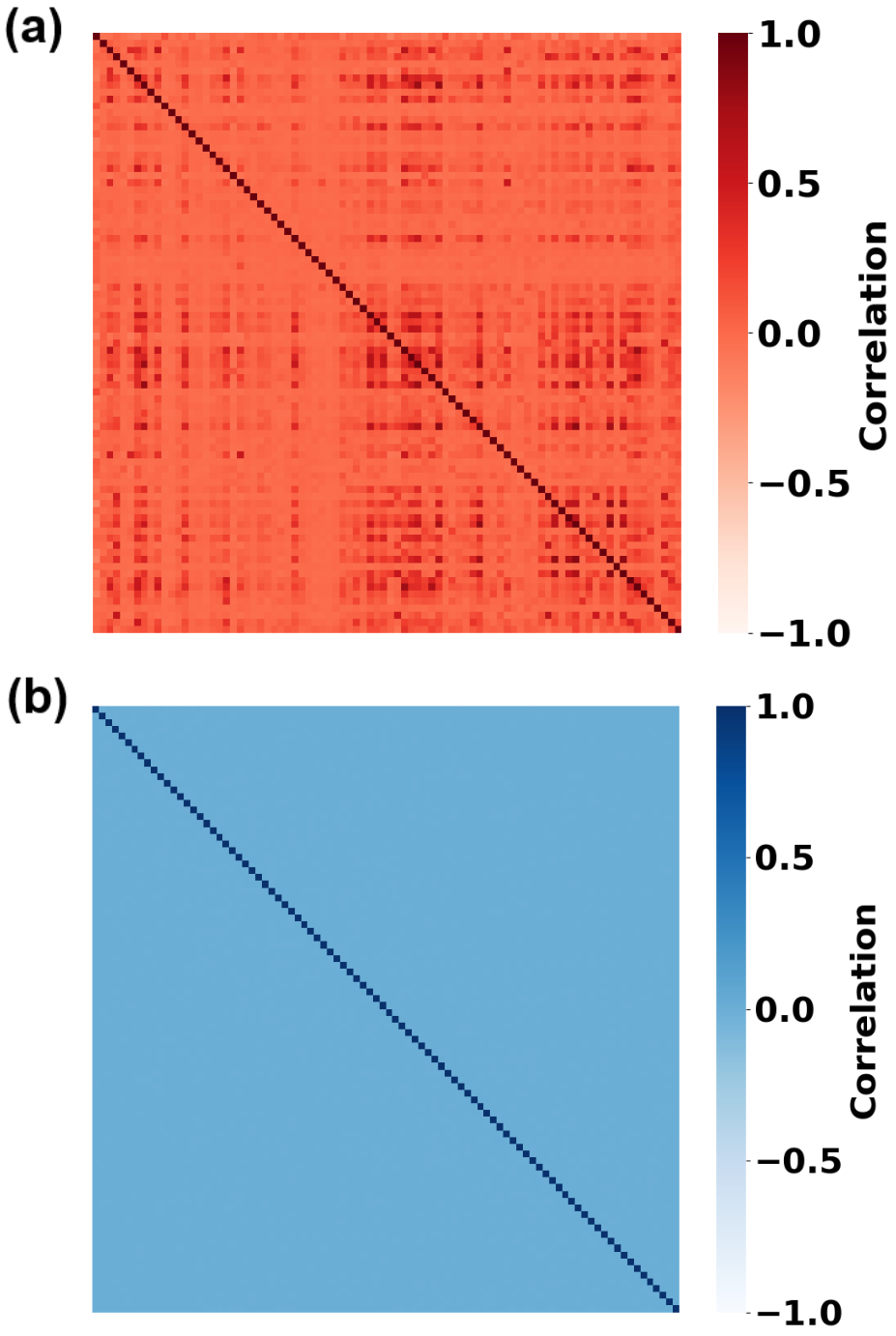

Fig. S7 Linear correlation analysis of features. Pearson correlation heatmaps of all components after dimensionality reduction of the XRD feature matrix using (a) NMF and (b) kPCA.

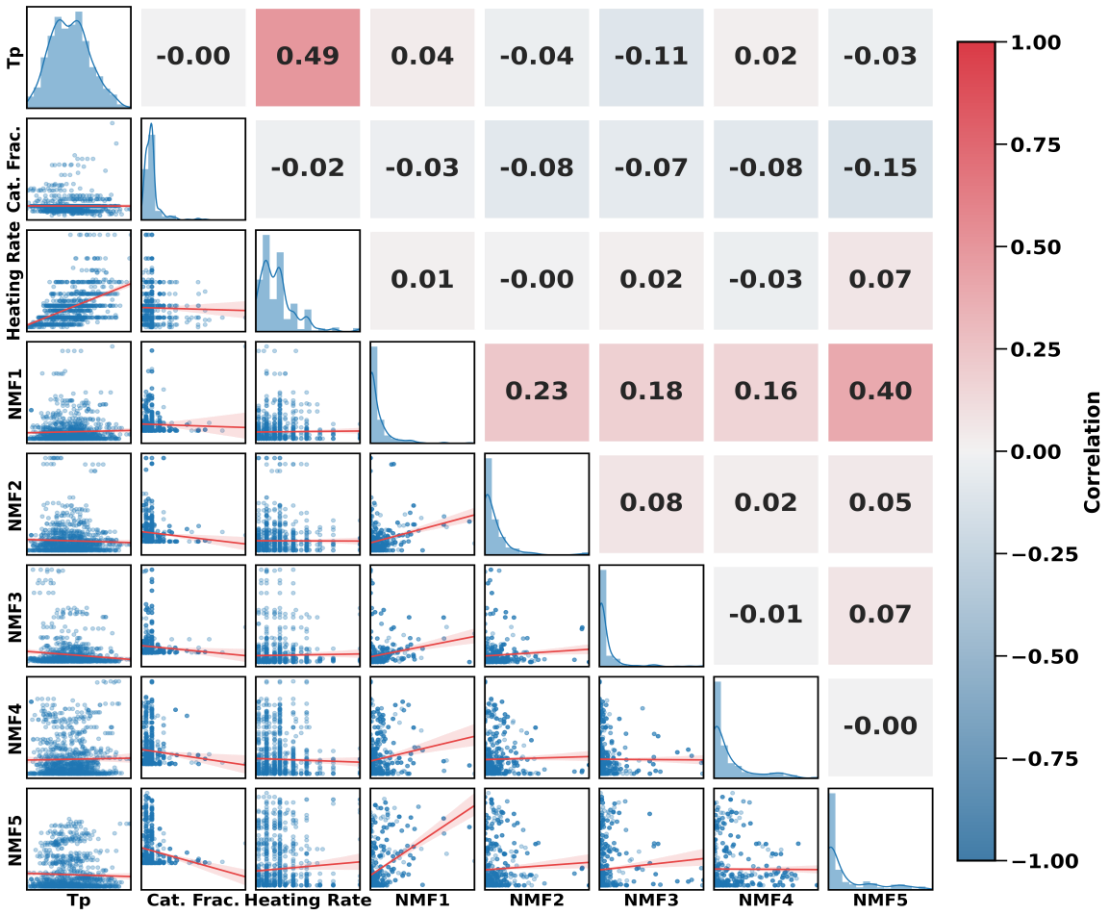

Fig. S8 Relationships among local NMF features, environmental variables, and the *Tp*. Correspond to EXP XRD data. Variables are ordered as follows from top to bottom on the left and left to right along the bottom: *Tp*, Cat. Frac., Heating Rate, NMF1, NMF2, NMF3, NMF4, NMF5. Each panel includes marginal distribution histograms on the diagonal, pairwise scatter plots with linear regression fits in the lower triangle, and a Pearson correlation heatmap in the upper triangle.

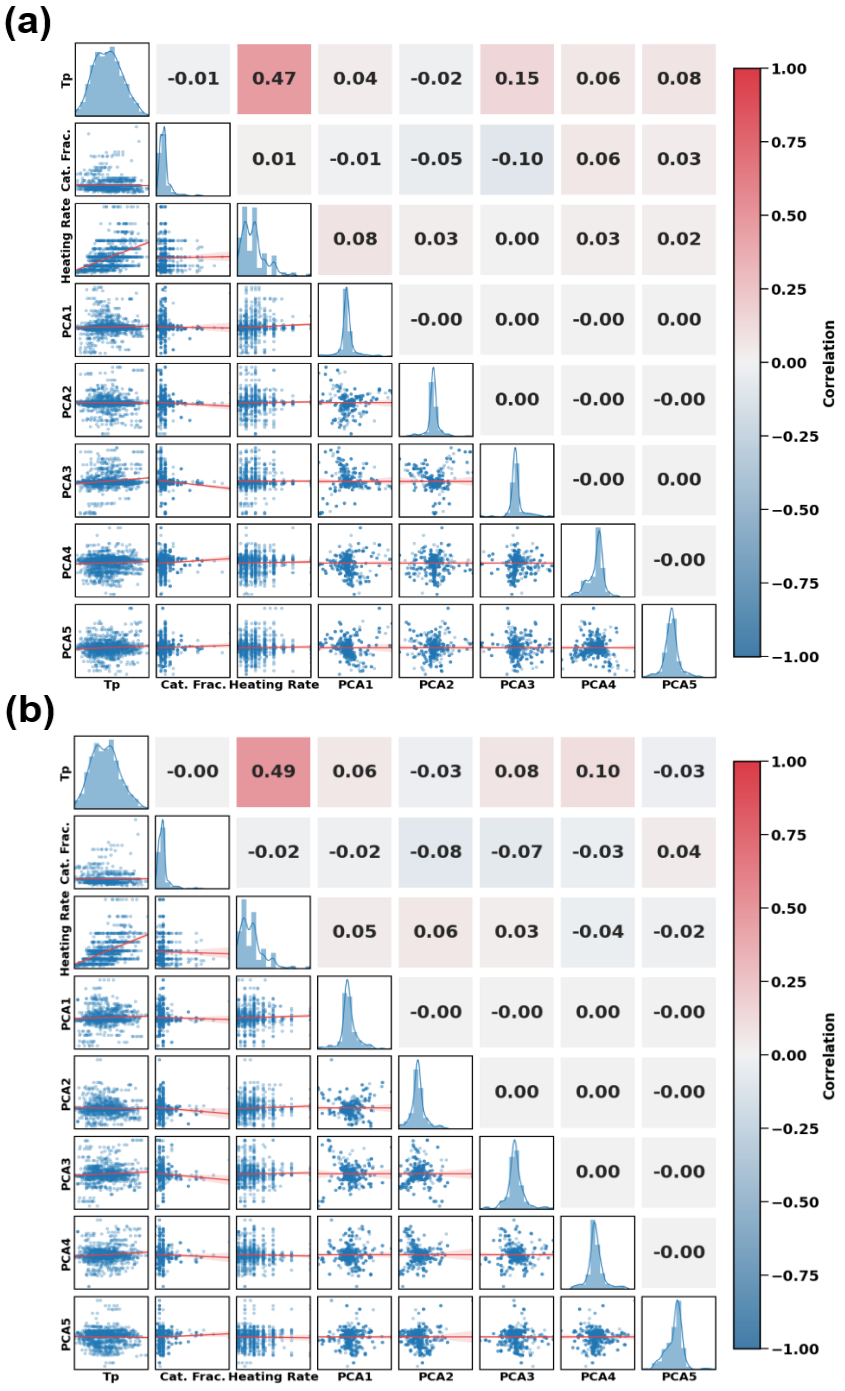

Fig. S9 Symmetric matrix plots illustrating the relationships among locally kPCA features, environmental variables, and the Tp: (a) derived from ICDD XRD data and (b) derived from EXP XRD data. The figures include marginal distribution histograms along the diagonal, pairwise scatter plots with linear regression fits in the lower triangle, and Pearson correlation heatmaps in the upper triangle.

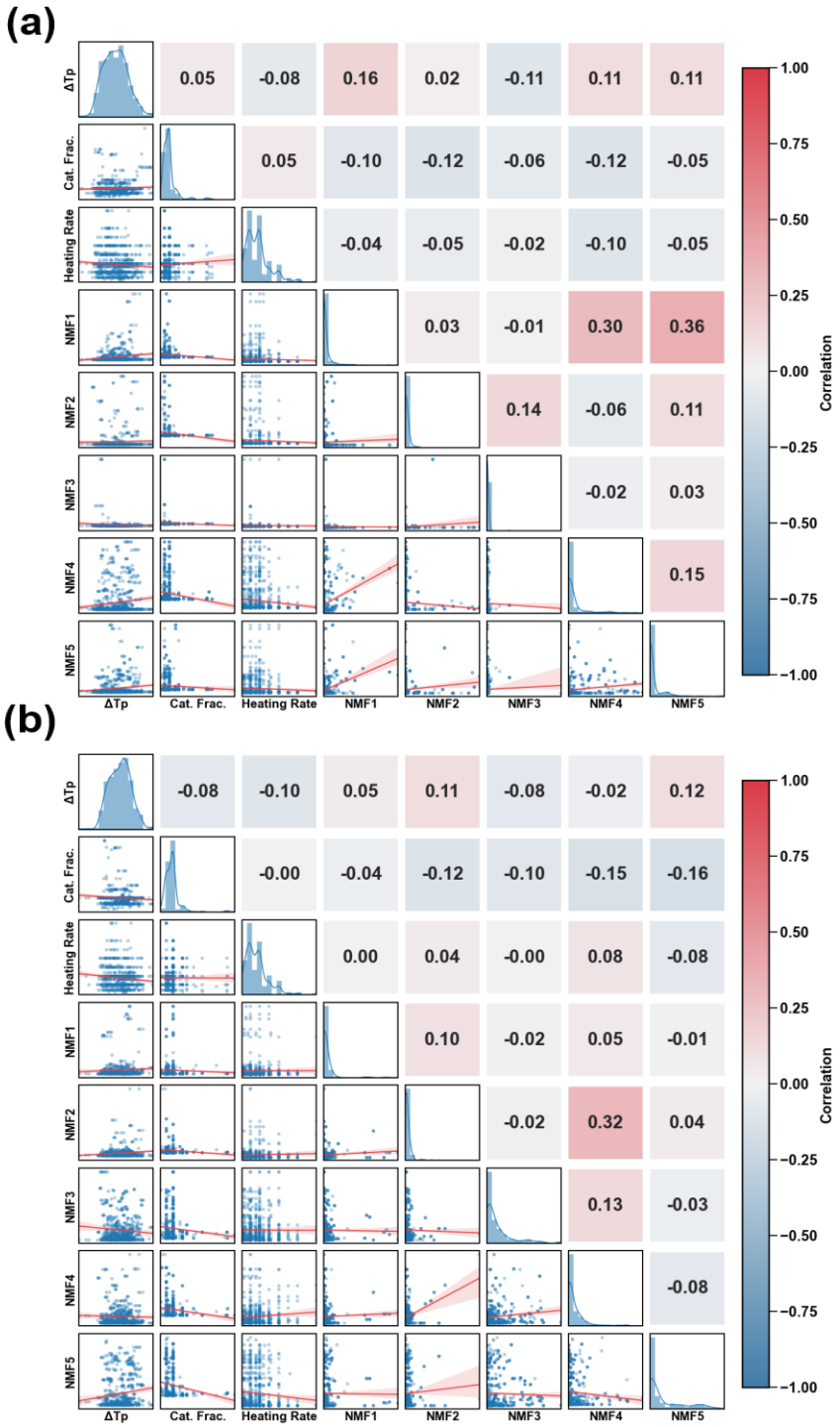

Fig. S10 Symmetric matrix plot showing the relationships between the locally NMF features, environmental variables, and the ΔTp: (a) derived from ICDD XRD data and (b) derived from EXP XRD data. The figure includes: marginal distribution histograms along the diagonal, pairwise scatter plots with linear regression fits in the lower triangle, and a Pearson correlation heatmap in the upper triangle.

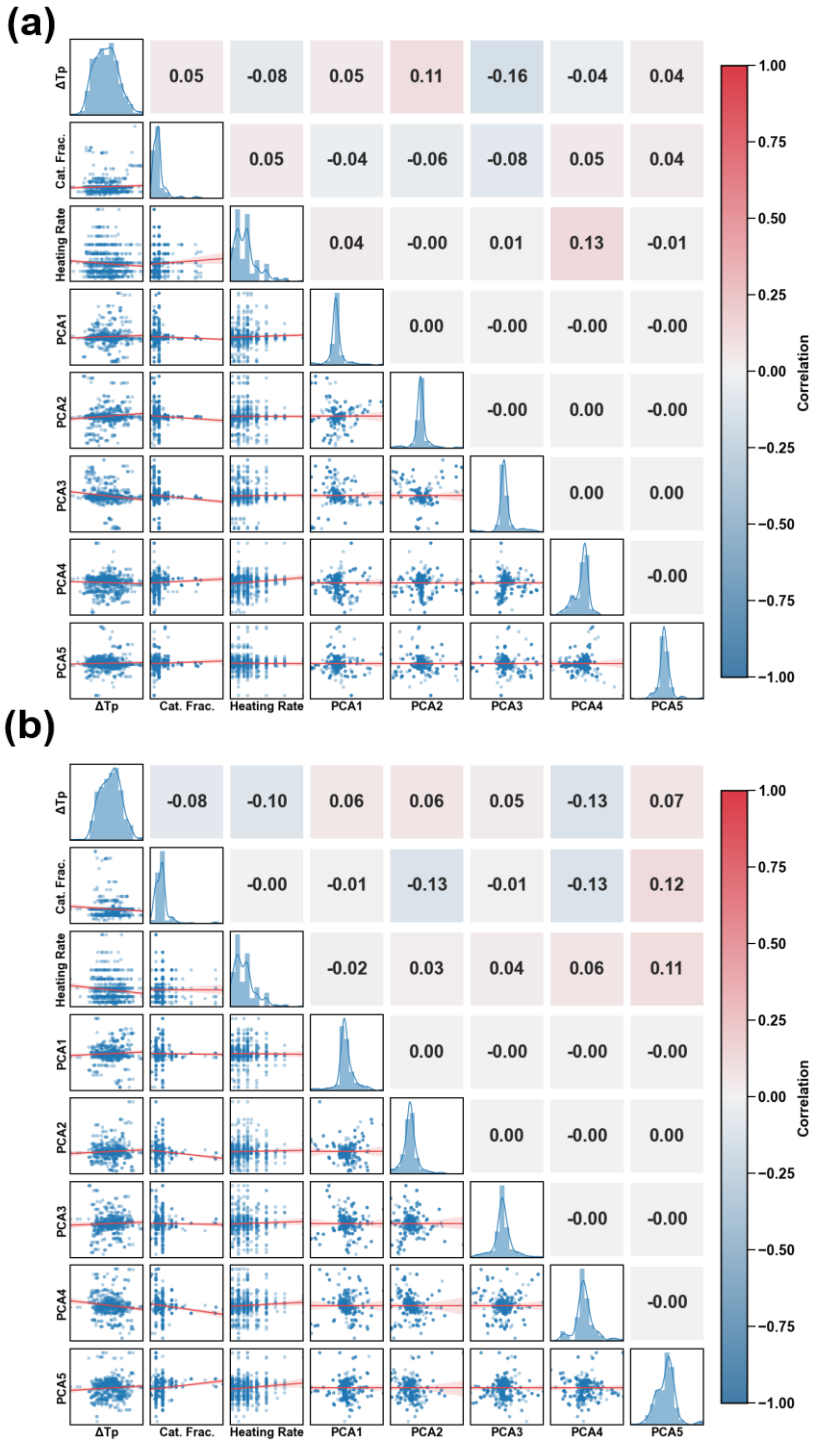

Fig. S11 Symmetric matrix plot showing the relationships between the locally kPCA features, environmental variables, and the ΔTp: (a) derived from ICDD XRD data and (b) derived from EXP XRD data. The figure includes: marginal distribution histograms along the diagonal, pairwise scatter plots with linear regression fits in the lower triangle, and a Pearson correlation heatmap in the upper triangle.

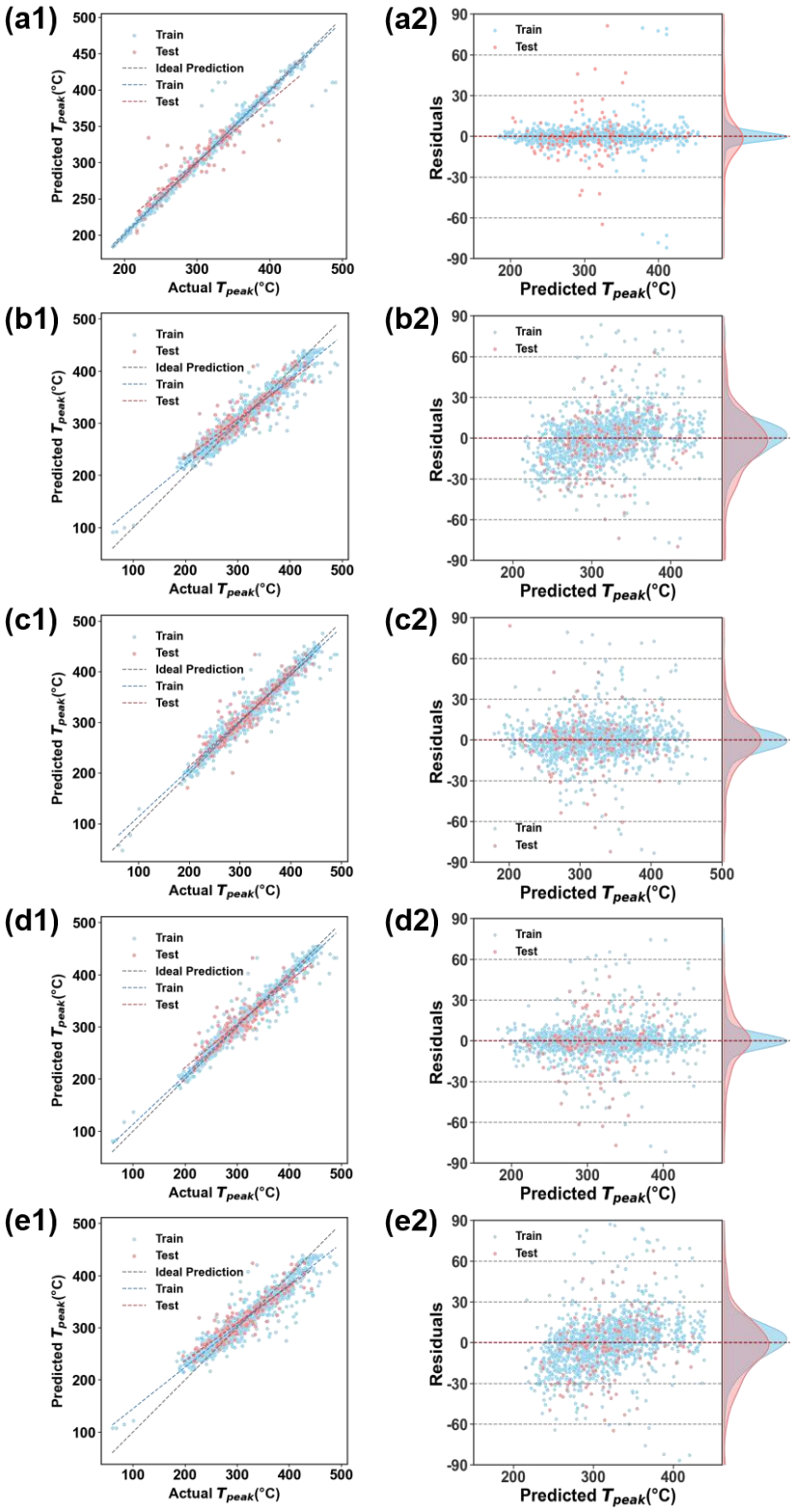

Fig. S12 Performance of non-optimal Tp prediction models constructed using ICDD XRD data. Each subplot includes a scatter plot of predicted vs. actual values (left) and residual distribution (right), corresponding to: (a) NMF+XGB, (b) NMF+RF, (c) kPCA+XGB, (d) kPCA+LGBM, and (e) kPCA+RF.

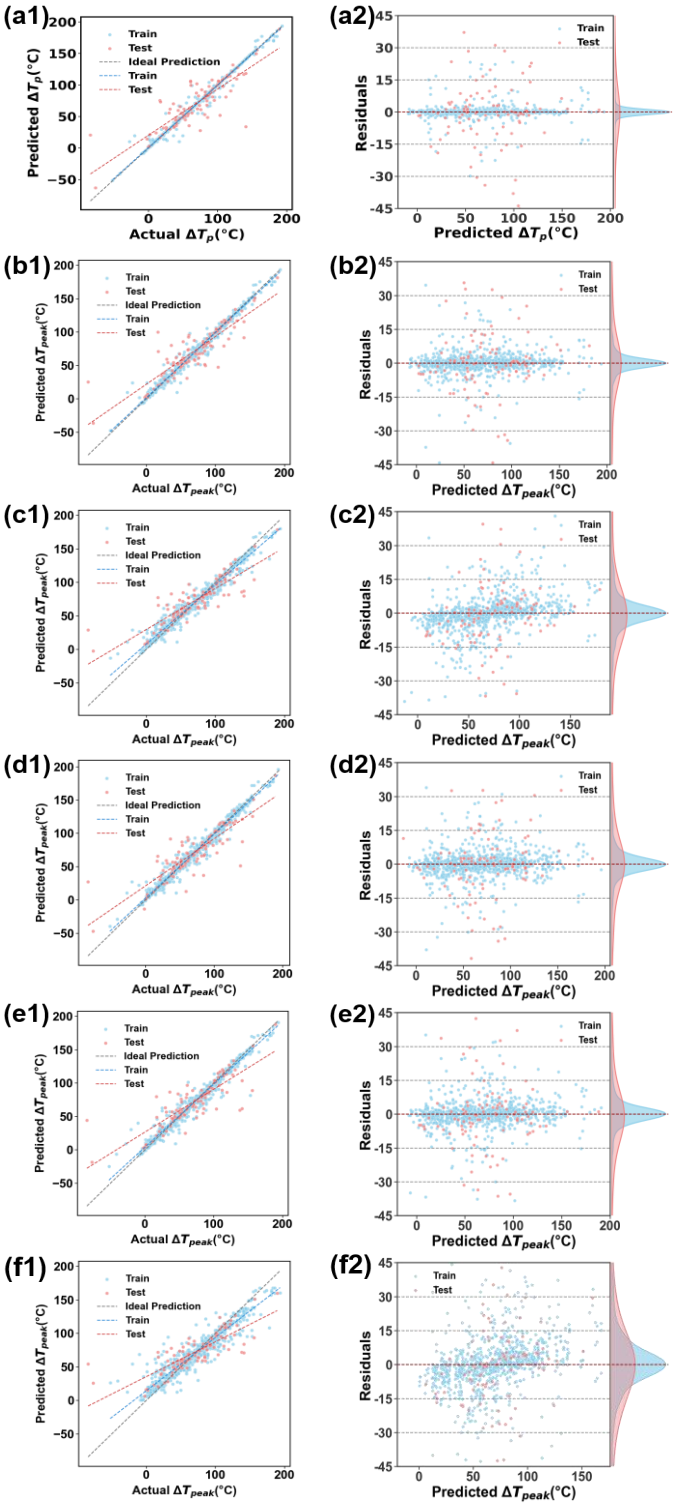

Fig. S13 Performance of non-optimal ΔTp prediction models constructed using ICDD XRD data. Each subplot includes a scatter plot of predicted vs. actual values (left) and residual distribution (right), corresponding to: (a) NMF+XGB, (b) NMF+LGBM, (c) NMF+RF, (d) kPCA+XGB, (e) kPCA+LGBM, (f) kPCA+RF.

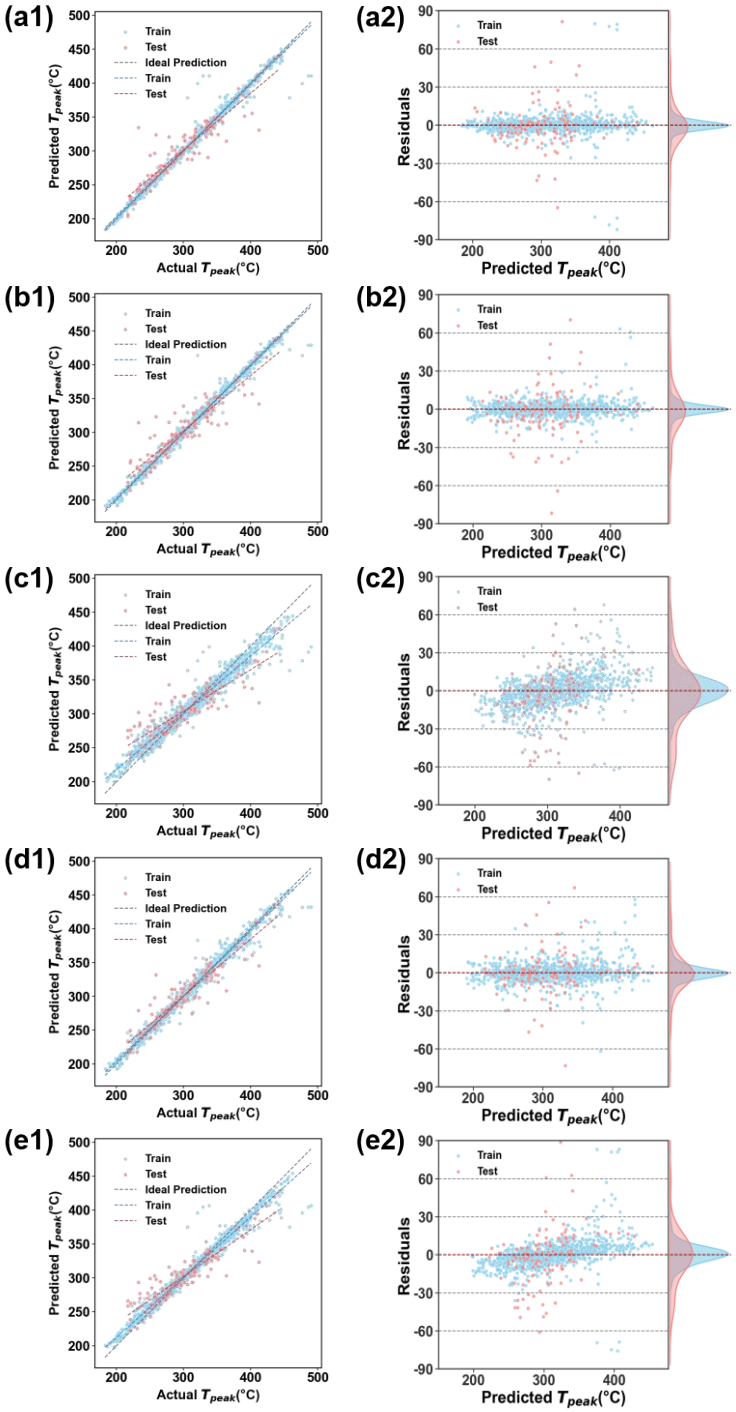

Fig. S14 Performance of non-optimal Tp prediction models constructed using EXP XRD data. Each subplot includes a scatter plot of predicted vs. actual values (left) and residual distribution (right), corresponding to: (a) NMF+XGB, (b) NMF+LGBM, (c) NMF+RF, (d) kPCA+LGBM, (e) kPCA+RF.

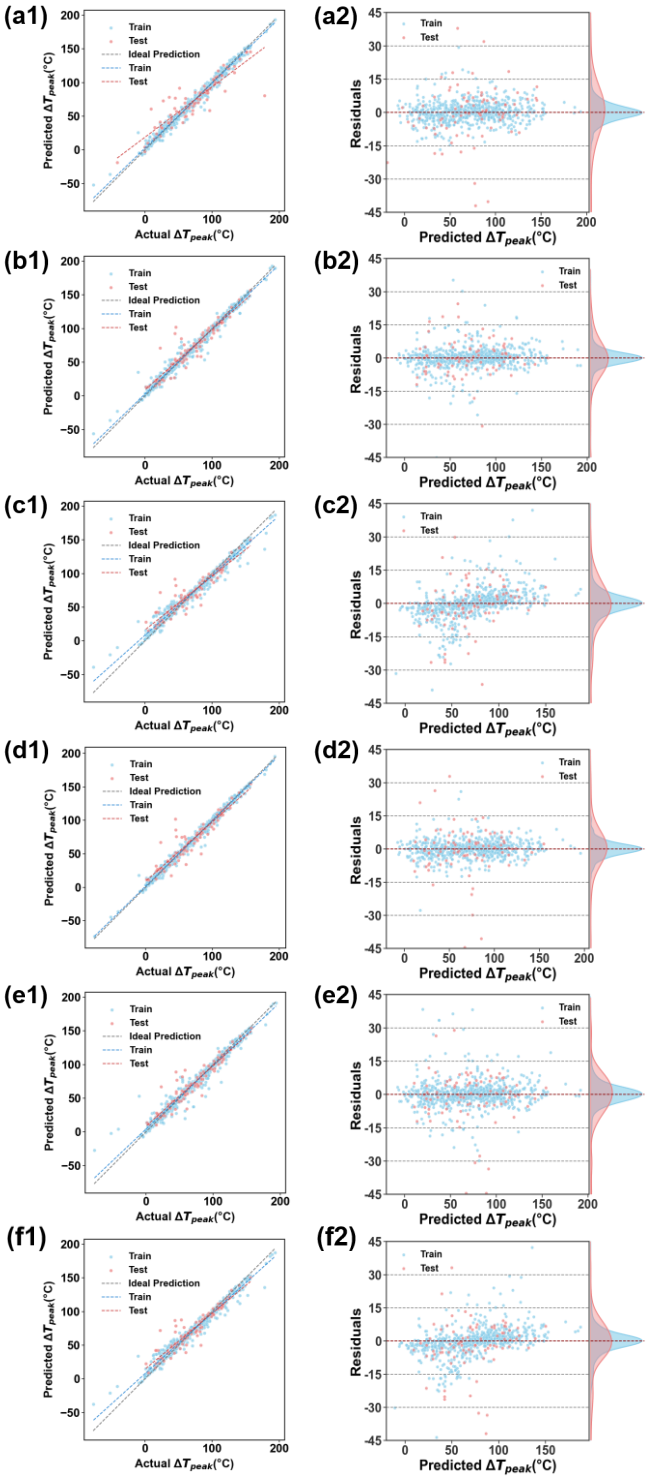

Fig. S15 Performance of non-optimal ΔTp prediction models constructed using EXP XRD data. Each subplot includes a scatter plot of predicted vs. actual values (left) and residual distribution (right), corresponding to: (a) NMF+XGB, (b) NMF+LGBM, (c) NMF+RF, (d) kPCA+XGB, (e) kPCA+LGBM, (f) kPCA+RF.

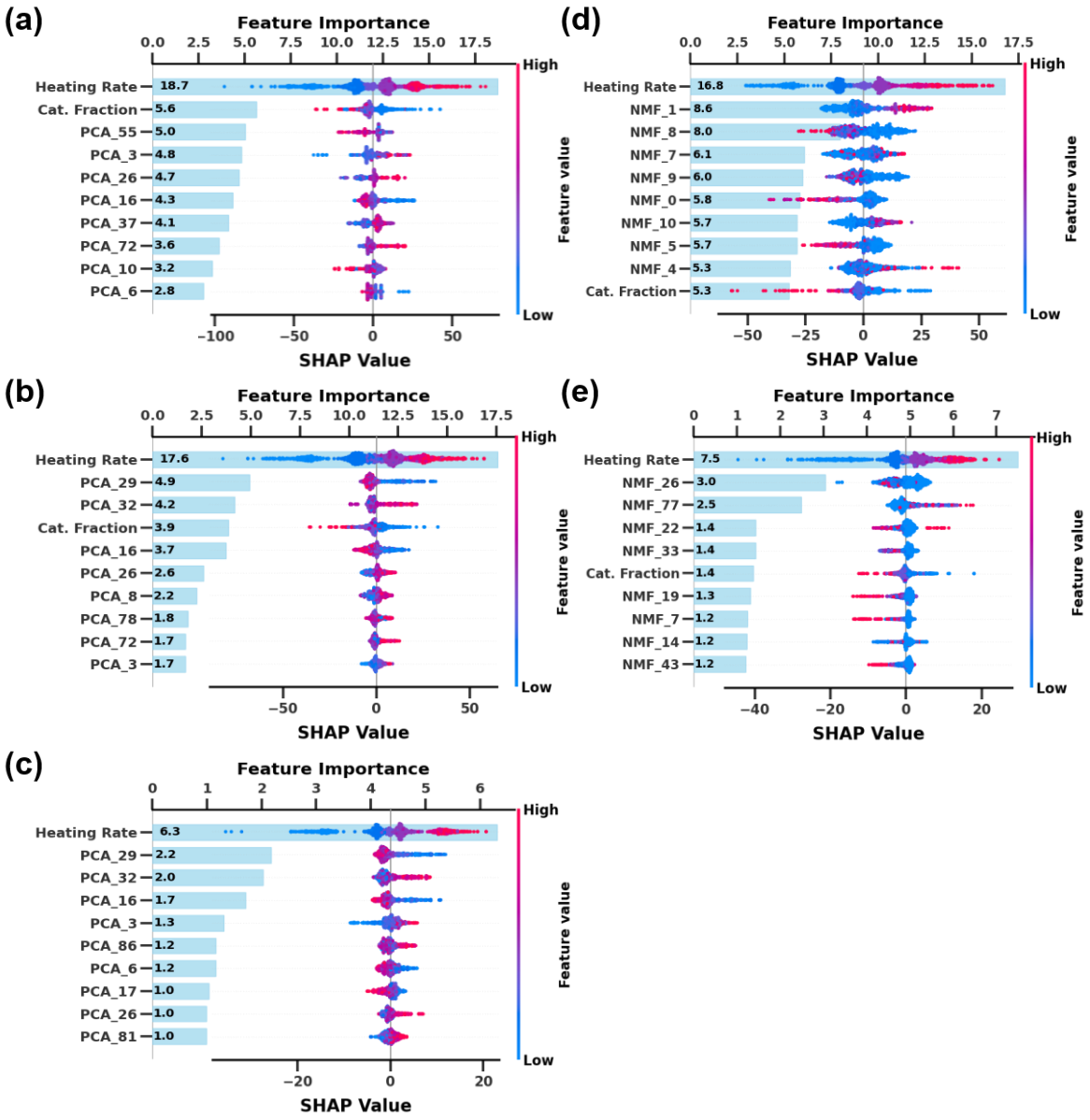

Fig. S16 Feature importance (bar plots) and SHAP analysis (dot plots) of suboptimal Tp models constructed using ICDD XRD data:(a) kPCA+XGB, (b) kPCA+LGBM, (c) kPCA+RF, (d) NMF+XGB, (e) NMF+RF.

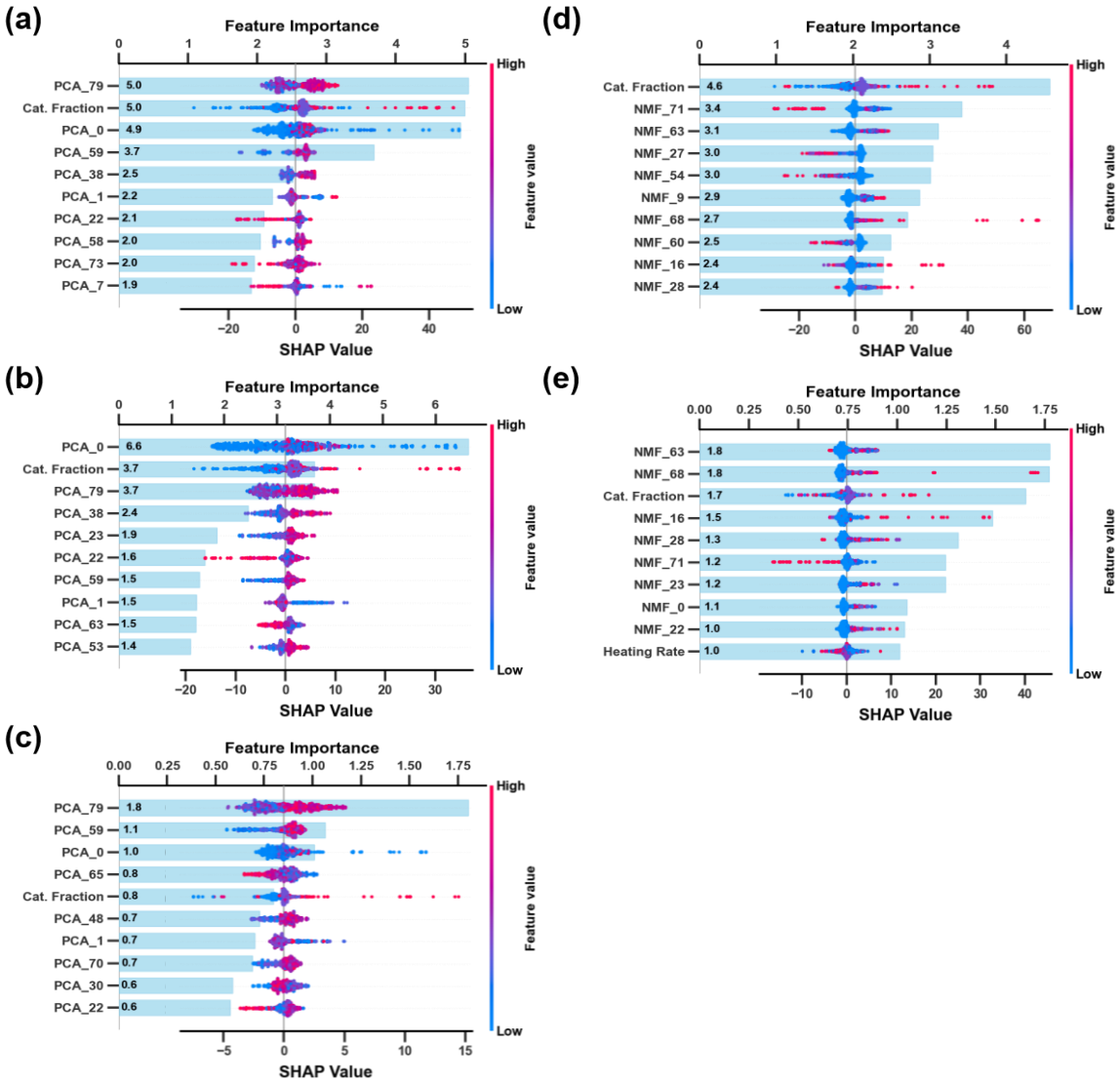

Fig. S17 Feature importance (bar plots) and SHAP analysis (dot plots) for ΔTp prediction models constructed using ICDD XRD data: (a) kPCA+XGB, (b) kPCA+LGBM, (c) kPCA+RF (d) NMF+LGBM, (e) NMF+RF,.

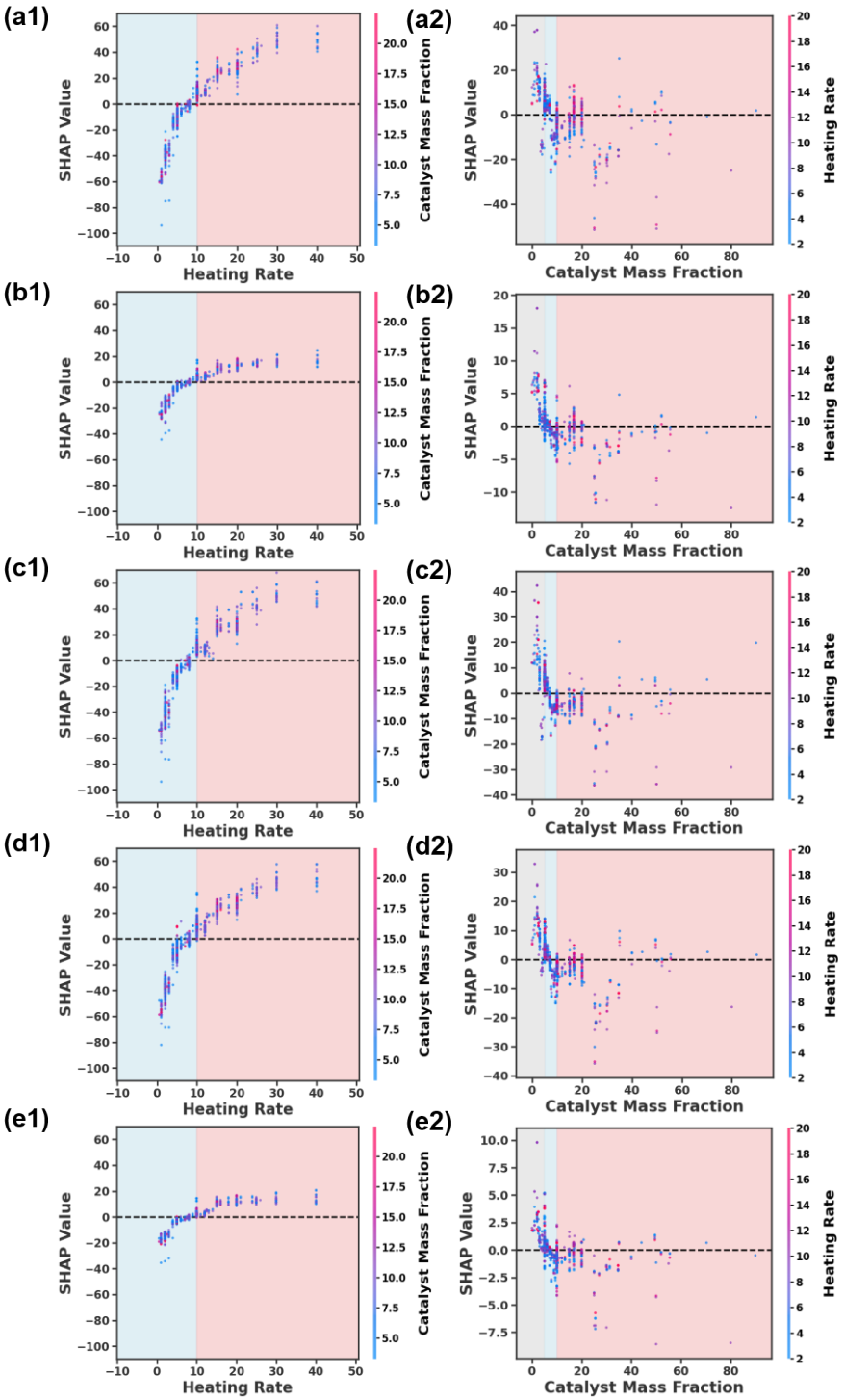

Fig. S18 Univariate SHAP analysis of non-optimal Tp prediction models constructed using ICDD XRD data. Each subplot includes SHAP dependence plots for heating rate (left) and catalyst loading (right), corresponding to:(a) NMF+XGB, (b) NMF+RF, (c) kPCA+XGB, (d) kPCA+LGBM, (e) kPCA+RF.

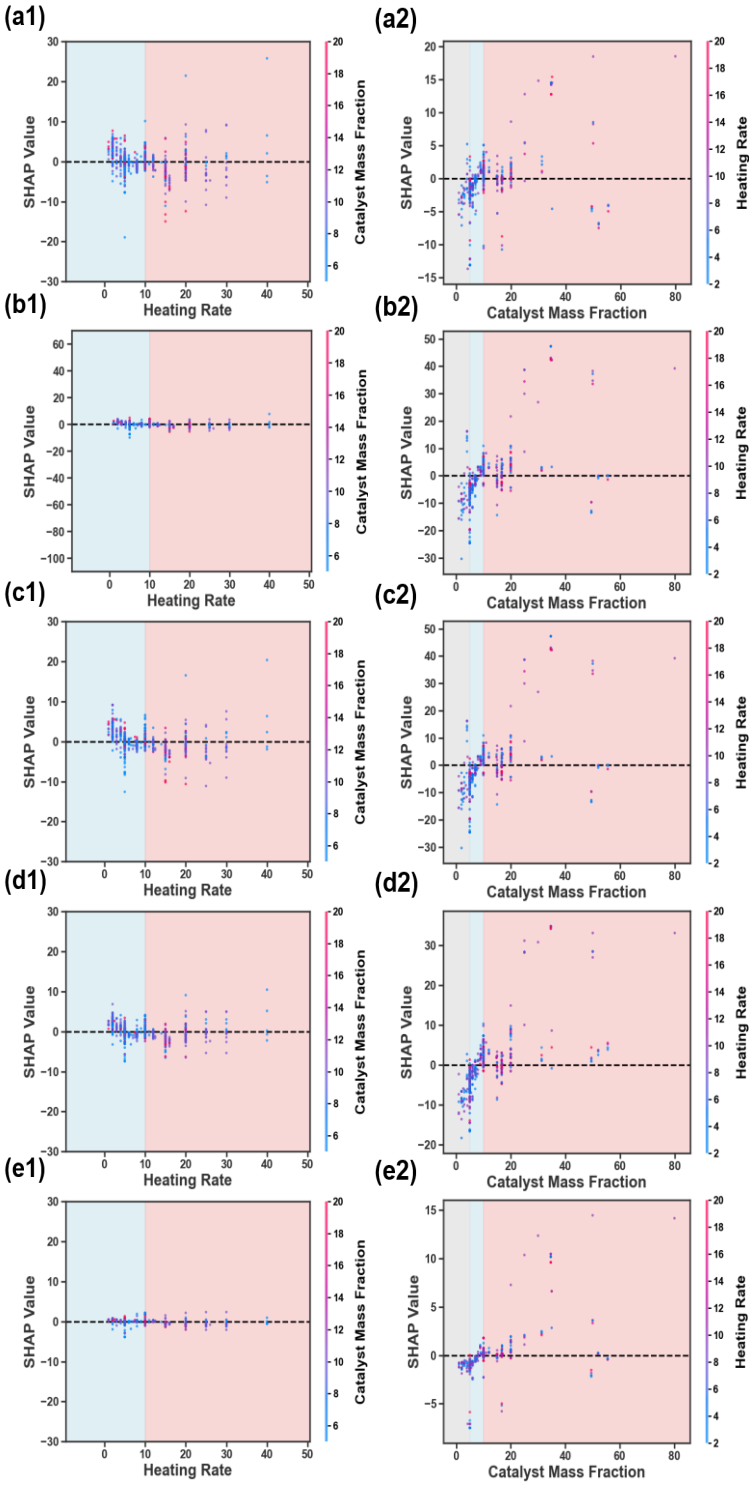

Fig. S19 Univariate SHAP analysis of ΔTp prediction models constructed using ICDD XRD data. Each subplot includes SHAP dependence plots for heating rate (left) and catalyst loading (right), corresponding to:(a) NMF+LGBM, (b) NMF+RF, (c) kPCA+XGB, (d) kPCA+LGBM, (e) kPCA+RF.

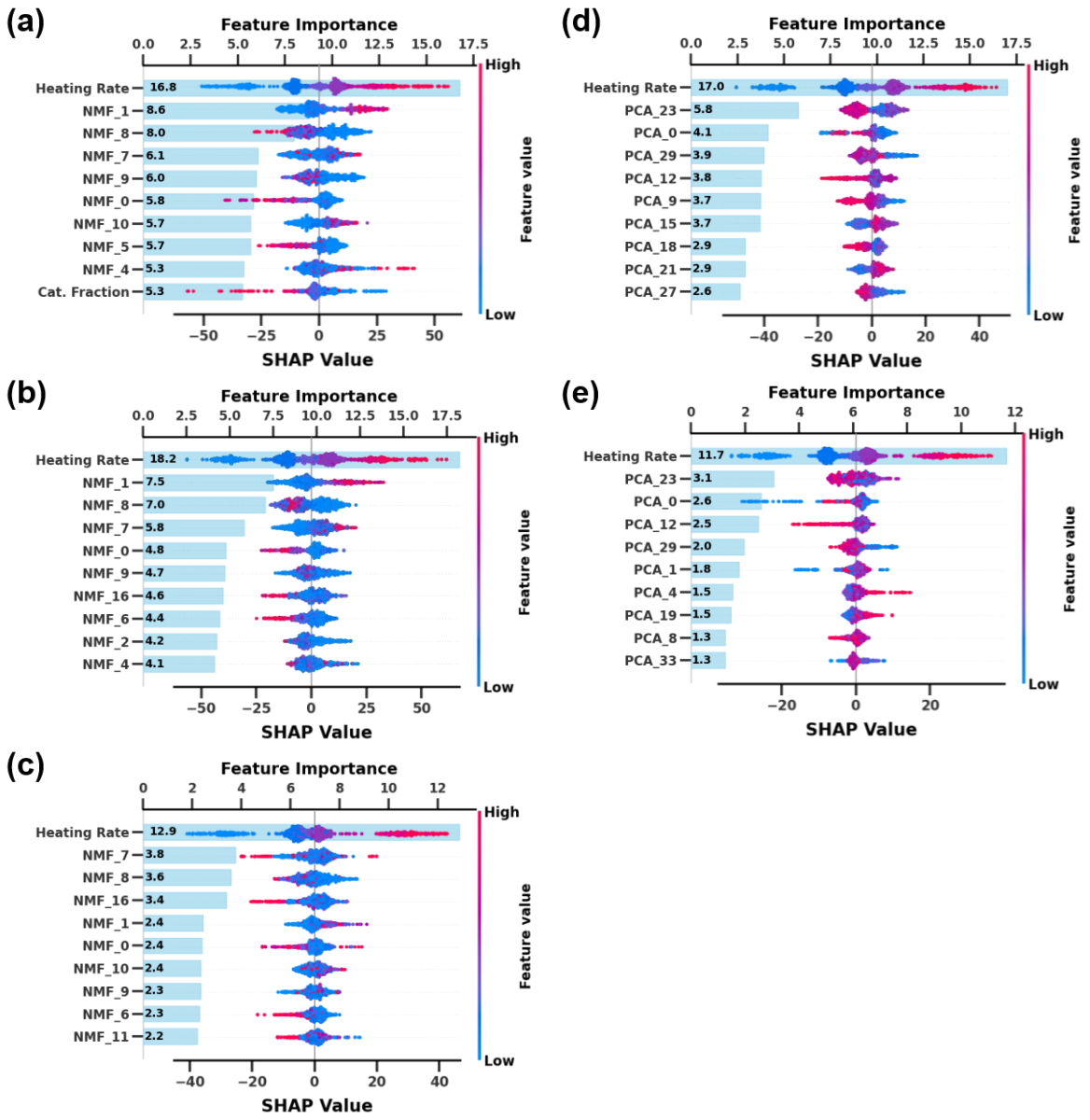

Fig. S20 Feature importance (bar plots) and SHAP analysis (dot plots) of Tp prediction models based on EXP XRD data, constructed using: (a) NMF+XGB, (b) NMF+LGBM, (c) NMF+RF, (d) kPCA+LGBM, (e) kPCA+RF.

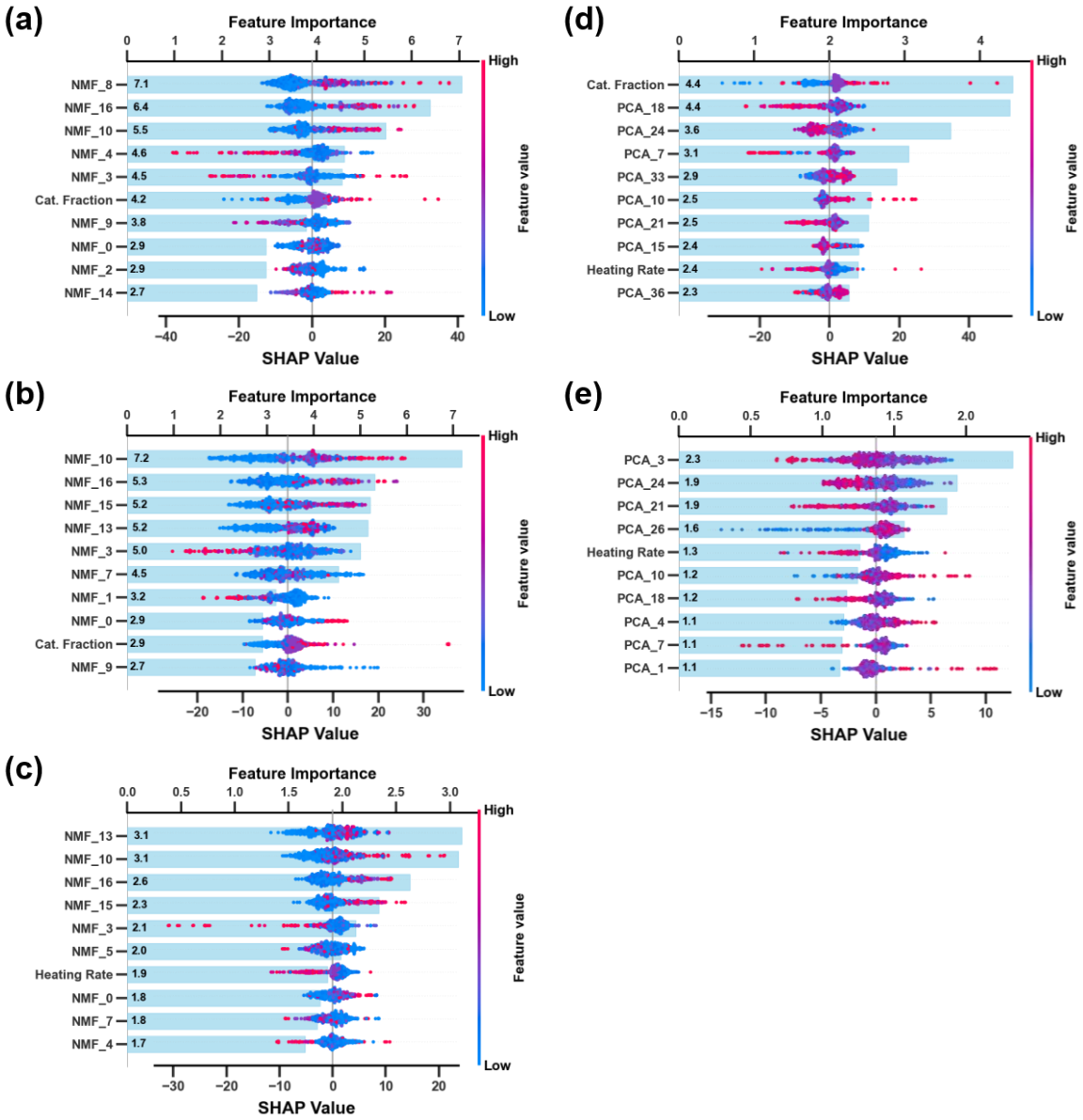

Fig. S21 Feature importance (bar plots) and SHAP analysis (dot plots) of ΔTp prediction models based on EXP XRD data, constructed using: (a) NMF+XGB, (b) NMF+LGBM, (c) NMF+RF, (d) kPCA+XGB, (e) kPCA+RF.

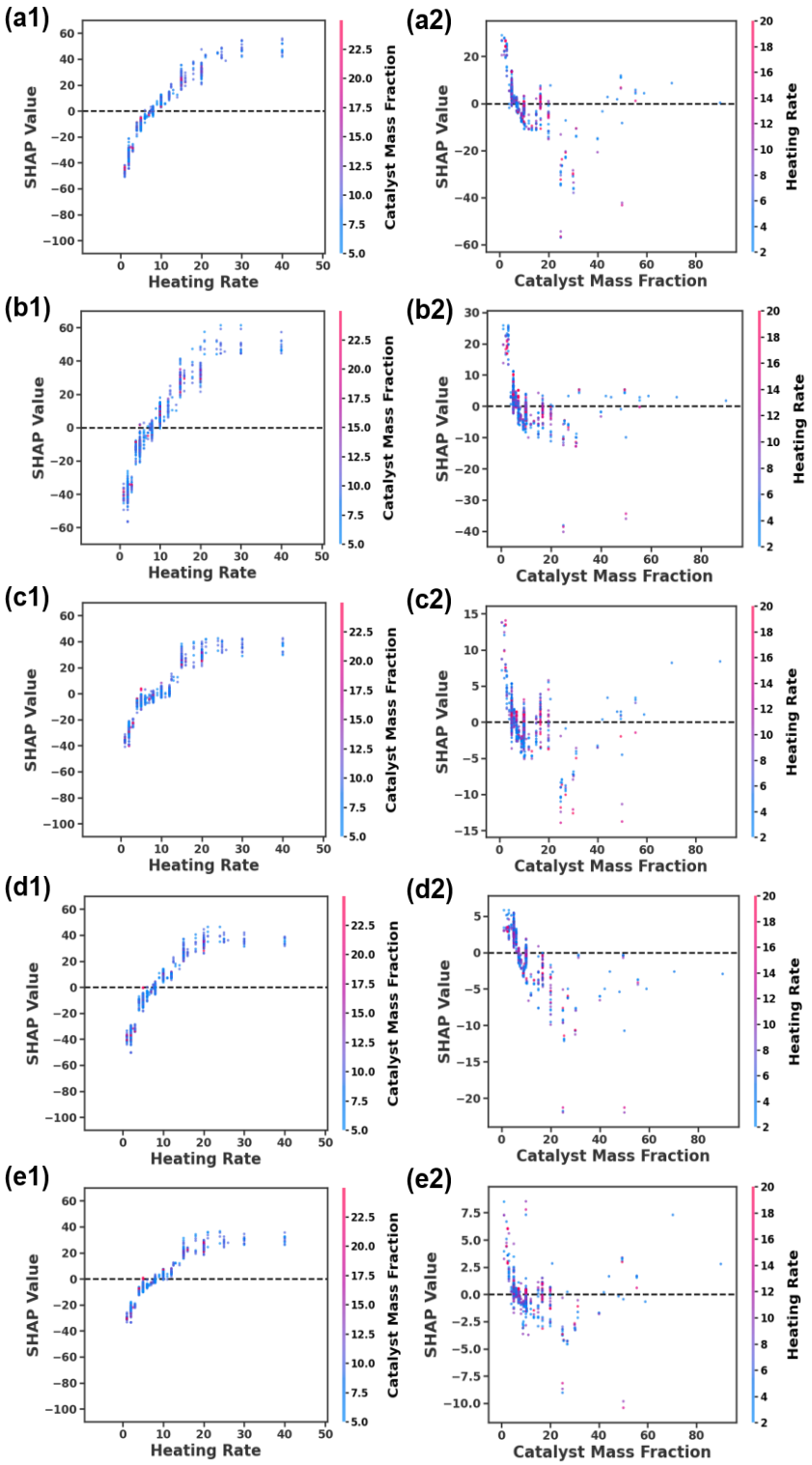

Fig. S22 Univariate SHAP analysis of Tp prediction models based on EXP XRD data. Each subfigure includes heating rate analysis (left) and catalyst loading analysis (right), corresponding to: (a) NMF+XGB, (b) NMF+LGBM, (c) NMF+RF, (d) kPCA+LGBM, (e) kPCA+RF.

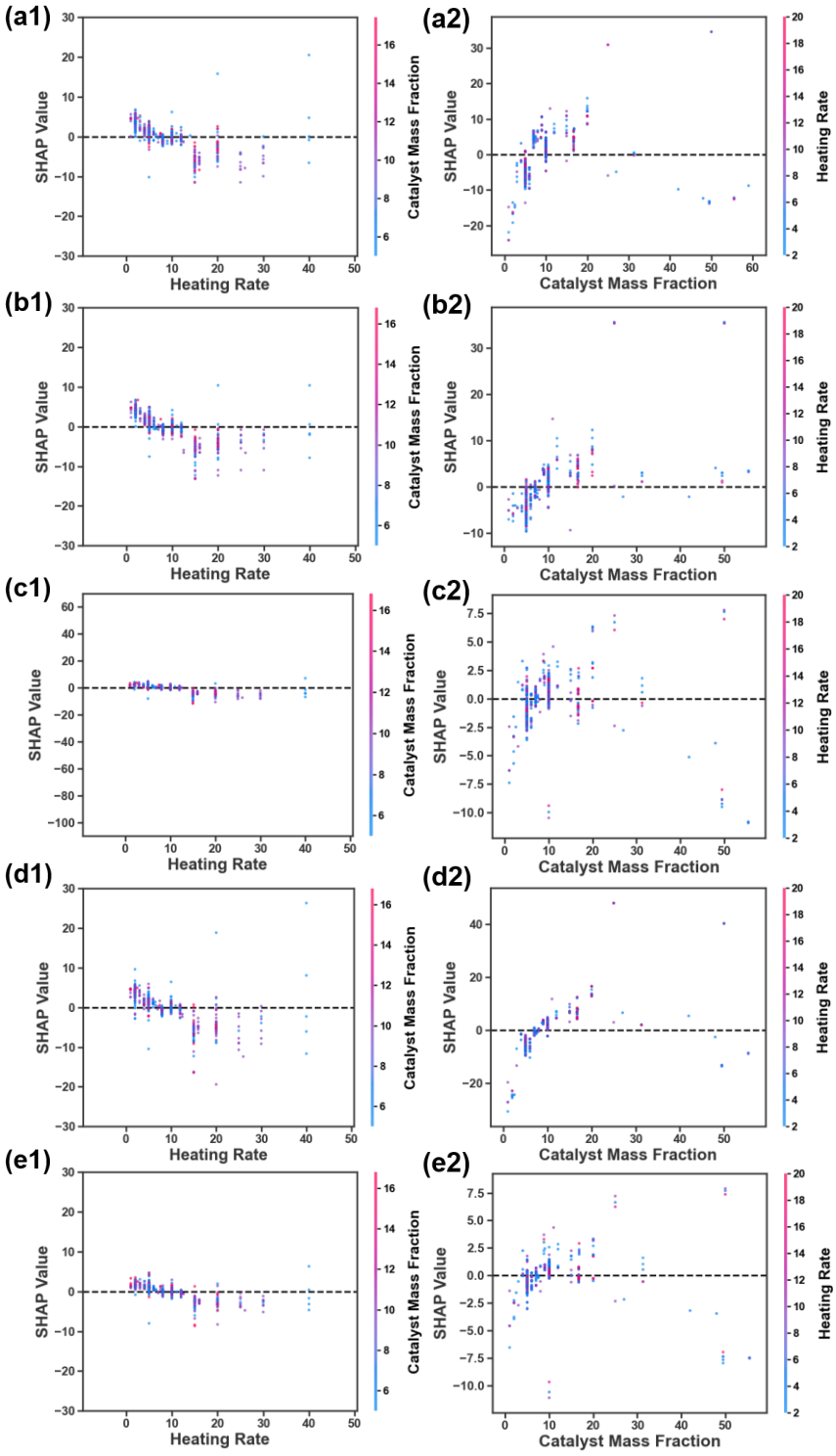

Fig. S23 Univariate SHAP analysis of ΔTp prediction models based on experimental XRD data. Each subfigure includes heating rate analysis (left) and catalyst loading analysis (right), corresponding to: (a) NMF+XGB, (b) NMF+LGBM, (c) NMF+RF, (d) kPCA+XGB, (e) kPCA+LGBM, and (f) kPCA+RF.

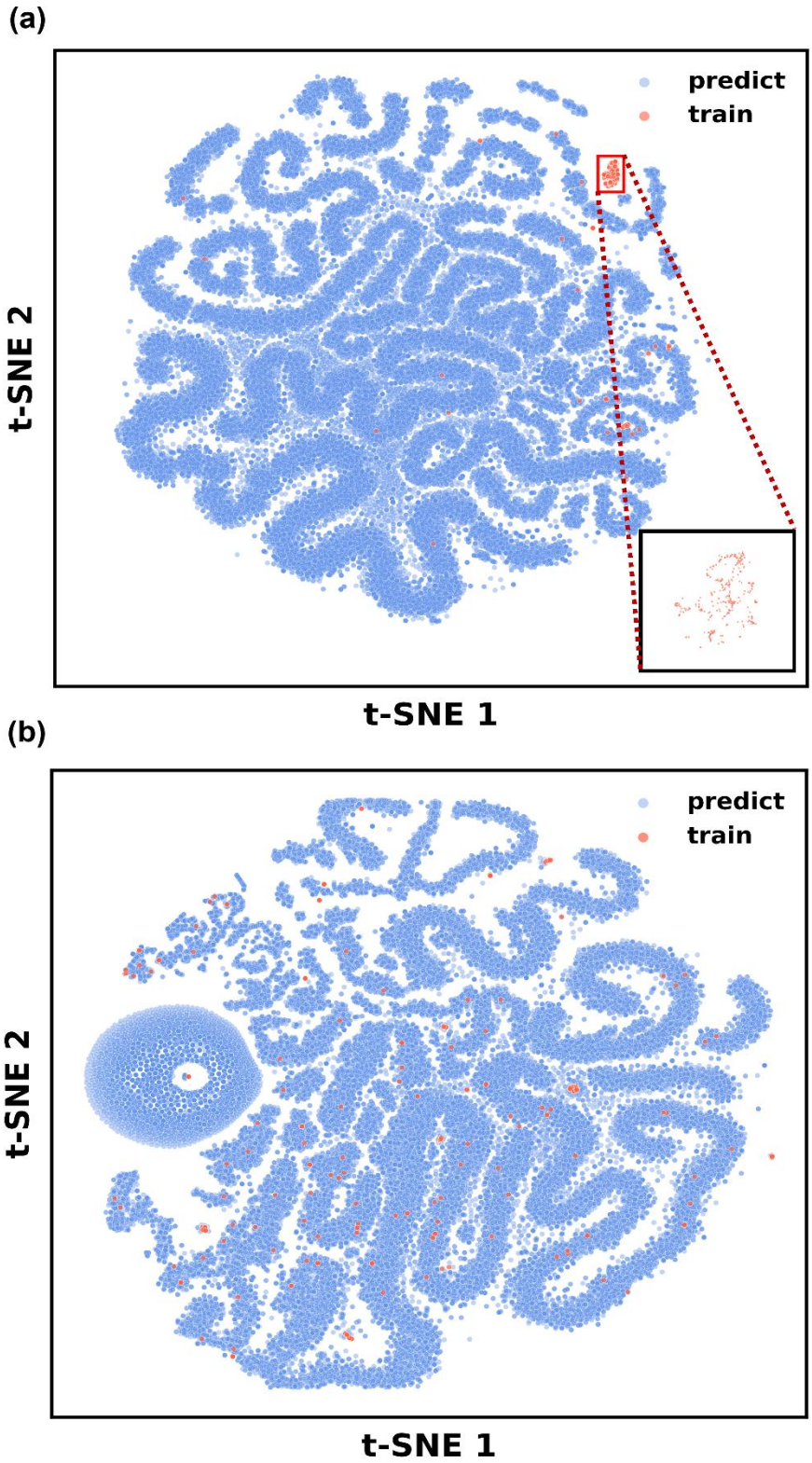

Fig. S24 (a) Distribution of the prediction set (blue), constructed by augmenting ICDD XRD with simulated experimental factors (such as grain refinement, etc.), and the training set (red) based on EXP XRD. The bottom left shows a magnified view of the region highlighted by the red box. (b) Distribution of the training set (red) using ICDD XRD as features, and the corresponding prediction set (blue).

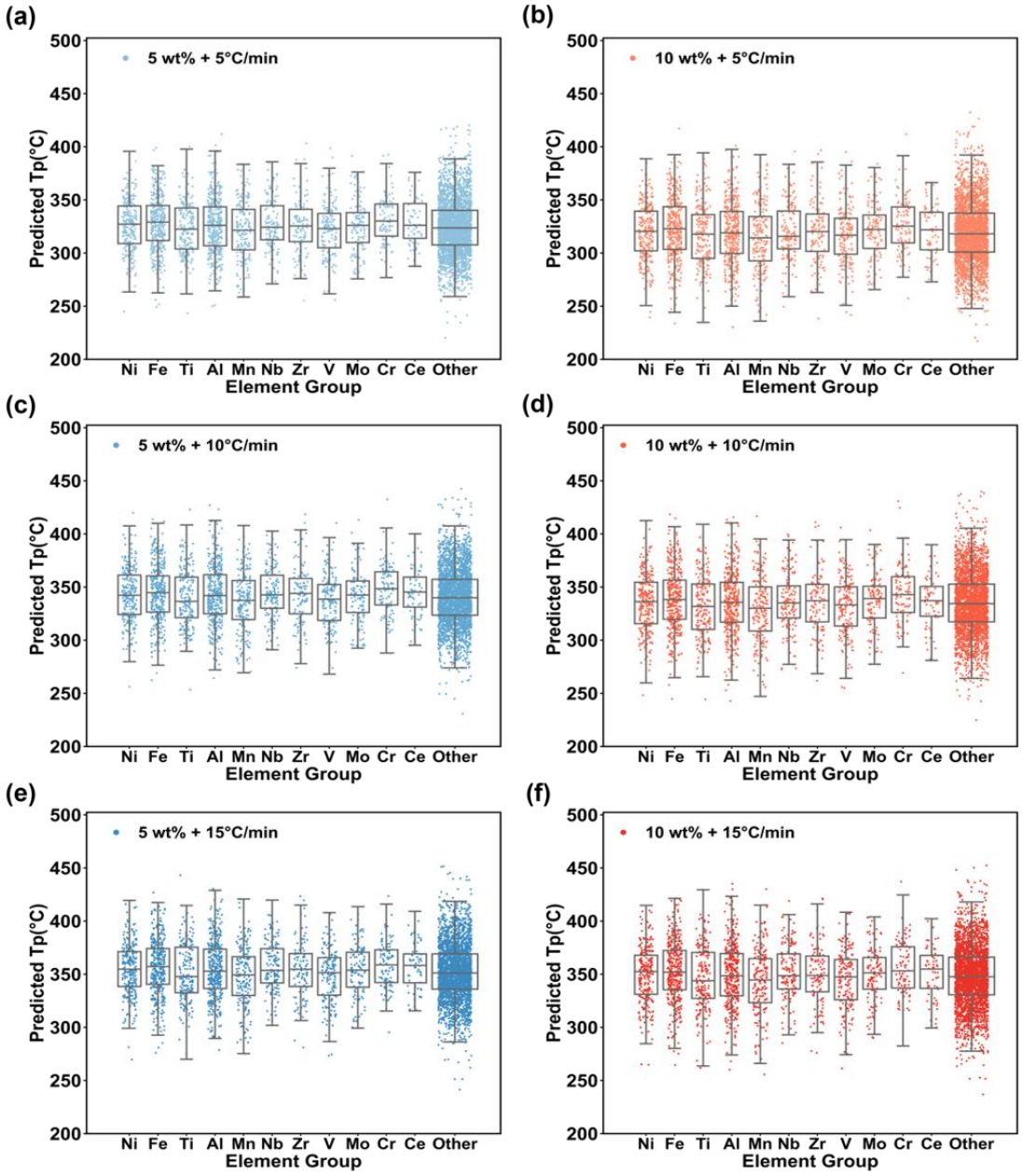

Fig. S25 Predicted results for 5,000 materials under 6 common environmental condition combinations.

Supplementary Tables

Table S1 MAE of Different Interpolation Methods on ICDD XRD Data at Various Step Sizes.

step	linear	nearest	cubic
0.01	2.94E-21	2.52E-07	6.68E-10
0.02	4.33E-10	5.04E-07	1.47E-07
0.03	0.013	0.058	0.004
0.04	0.015	0.087	0.002
0.05	0.032	0.104	0.008
0.06	0.039	0.116	0.012
0.07	0.060	0.148	0.024
0.08	0.069	0.173	0.037
0.09	0.093	0.194	0.056
0.1	0.102	0.205	0.071

Table S2 MAE of Different Interpolation Methods on EXP XRD Data at Various Step Sizes.

step	linear	quadratic	cubic
0.1	0.019	0.020	0.020
0.09	0.018	0.019	0.019
0.08	0.017	0.018	0.018
0.07	0.016	0.017	0.017
0.06	0.015	0.015	0.016
0.05	0.014	0.014	0.014
0.04	0.012	0.012	0.012
0.03	0.011	0.010	0.011
0.02	0.008	0.008	0.008
0.01	0.005	0.005	0.005

Table S3 Fit Metrics, Error Metrics, and Residual Analysis Metrics for Training Set, Test Set, and 10-Fold-Cross-Validation after Hyperparameter Tuning for Various Dimensionality Reduction Methods and Model Combinations Based on ICDD XRD Data.

Model	Metric	Tp(°C)			ΔTp(°C)		
		Train	Test	10 Fold	Train	Test	10 Fold
NMF+XGB	R^2	0.961	0.818	0.823	0.984	0.789	0.775
	RMSE	11.64	22.89	24.15	5.01	15.8	18.9
	MAE	6.25	14.92	14.79	1.44	11.24	11.45
	Mean resid	0	-4.28	—	0	0.08	—
	Std Dev resid	11.63	22.48	—	5.01	15.8	—
	Skewn resid	0.17	-1.37	—	-5.04	-0.72	—
NMF+LGBM	R^2	0.968	0.848	0.822	0.984	0.762	0.752
	RMSE	10.64	20.9	24.29	5.03	21.26	19.25
	MAE	5.36	12.74	14.630	2.62	12.36	11.42
	Mean resid	0	-3.13	—	0	-1.87	—
	Std Dev resid	10.63	20.66	—	5.03	21.18	—
	Skewn resid	-0.04	-1.91	—	-0.92	-1.64	—
NMF+RF	R^2	0.91	0.808	0.775	0.959	0.718	0.777
	RMSE	17.71	23.5	27.5	8.05	23.15	18.4
	MAE	12.44	16.88	19.74	4.89	14.25	11.97
	Mean resid	0.47	-3.94	—	-0.29	-2.4	—
	Std Dev resid	17.71	23.16	—	8.05	23.02	—
	Skewn resid	0.71	-0.54	—	-0.52	-1.01	—
kPCA+XGB	R^2	0.946	0.831	0.804	0.975	0.741	0.75
	RMSE	13.82	22.06	26.1	6.3	22.2	19.61
	MAE	8.78	14.13	15.84	3.72	13.12	11.51
	Mean resid	0.000	-2.380	—	0	-0.59	—

	Std Dev resid	13.820	21.930	—	6.3	22.19	—
	Skewn resid	0.220	-0.850	—	-0.63	-0.6	—
kPCA+LGBM	R^2	0.959	0.840	0.802	0.966	0.735	0.769
	RMSE	12.000	21.400	25.74	7.340	22.46	18.78
	MAE	6.810	14.160	16.36	3.9	13.49	11.36
	Mean resid	0	-2.23	—	0.000	-1.320	—
	Std Dev resid	11.99	21.29	—	7.340	24.420	—
	Skewn resid	0.21	-1.29	—	-0.020	-0.750	—
kPCA+RF	R^2	0.888	0.789	0.746	0.886	0.684	0.735
	RMSE	19.78	24.6	29.3	13.47	24.73	20.23
	MAE	13.69	17.73	21.2	8.47	15.160	13.93
	Mean resid	0.43	-4.48	—	-0.21	-1.66	—
	Std Dev resid	19.78	24.19	—	13.47	26.67	—
	Skewn resid	0.65	-0.51	—	-0.75	-1.34	—

Table S4 Fitting metrics, error metrics, and residual analysis results for combinations of dimensionality reduction methods and regression models based on EXP XRD data, evaluated on the training set, test set, and 10-fold cross-validation after hyperparameter optimization.

Model	Metric	Tp(°C)			ΔTp(°C)		
		Train	Test	10 Fold	Train	Test	10 Fold
NMF+XGB	R^2	0.974	0.852	0.864	0.984	0.805	0.765
	RMSE	8.035	19.496	20.285	4.829	18.063	17.772
	MAE	3.517	11.134	11.662	3.28	10.761	10.077
	Mean resid	0	-1.23	—	0	-0.33	—
	Std Dev resid	8.03	19.46	—	4.83	18.06	—
	Skewn resid	0.19	-0.52	—	-0.19	1.29	—
NMF+LGBM	R^2	0.976	0.851	0.851	0.981	0.86	0.783
	RMSE	7.093	19.597	21.537	5.278	12.523	17.599
	MAE	3.555	12.814	13.766	3.049	7.302	9.89
	Mean_resid	0	-2.18	—	0	-1.63	—
	Std Dev_resid	7.09	19.47	—	5.28	12.42	—
	Skewn_resid	-2.42	-0.28	—	-0.5	-2.16	—
NMF+RF	R^2	0.937	0.728	0.784	0.968	0.859	0.785
	RMSE	14.266	26.424	26.076	6.883	13.025	17.726
	MAE	9.768	18.759	18.706	4.183	8.677	11.157
	Mean_resid	0.24	-3.25	—	-0.26	-2.11	—
	Std Dev_resid	14.26	26.22	—	6.88	12.85	—
	Skewn_resid	1.06	0.11	—	-0.28	-1.28	—
kPCA+XGB	R^2	0.975	0.88	0.88	0.985	0.865	0.814
	RMSE	7.593	17.551	18.89	3.857	12.767	16.329
	MAE	2.923	9.935	10.919	2.625	7.73	9.118
	Mean_resid	0	-0.99	—	0	-1.22	—

	Std Dev_resid	7.59	17.52	—	3.86	12.71	—
	Skewn_resid	0.02	0.8	—	0.25	-1.61	—
kPCA+LGBM	R^2	0.977	0.88	0.869	0.965	0.885	0.788
	RMSE	8.57	17.526	19.949	7.204	11.797	17.452
	MAE	4.993	10.976	12.149	3.938	7.738	9.767
	Mean_resid	0.06	-0.18	—	0.01	-1.44	—
	Std Dev_resid	8.57	17.53	—	7.2	11.71	—
	Skewn_resid	-0.51	0.17	—	-1.32	-1.35	—
kPCA+RF	R^2	0.963	0.809	0.82	0.97	0.865	0.807
	RMSE	10.873	22.159	23.7	6.647	12.751	16.741
	MAE	6.518	15.458	16.088	4.026	8.263	10.479
	Mean_resid	0.34	-1.48	—	-0.37	-2.58	—
	Std Dev_resid	10.87	22.11	—	6.64	12.49	—
	Skewn_resid	0.86	0.43	—	-0.47	-1.2	—

Table S5 Evaluation metrics of the ensemble models trained using Materials Project features.

Model	Metric	Tp		ΔTp	
		Train	Test	Train	Test
LGBM	R^2	0.9648	0.7286	0.9719	0.7678
	RMSE	4.2467	31.3495	0.4521	20.2632
	MAE	6.449	21.2871	4.0237	13.7961
RF	R^2	0.844	0.6485	0.8324	0.6857
	RMSE	23.2067	35.6792	16.9013	23.5751
	MAE	16.6152	26.4776	11.9954	17.5439
XGB	R^2	0.9792	0.7616	0.9892	0.81
	RMSE	8.4748	29.3825	4.2864	18.3265
	MAE	5.3128	19.5527	2.9031	12.812

Table S6 Evaluation metrics of the 1D-CNN models trained on ICDD-XRD and EXP-XRD datasets.

Model	Metric	Tp(°C)		ΔTp(°C)	
		Train	Test	Train	Test
ICDD	R^2	0.83	0.77	0.89	0.72
	MAE	15.36	16.23	8.34	14.32
EXP	R^2	0.82	0.74	0.86	0.71
	MAE	15.74	16.65	8.97	14.66

Table S7 Evaluation metrics for the ablation configuration with XRD-derived features removed

Model	Split	R^2	MAE	RMSE
XGB	Train	0.448	33.544	42.182
	Test	0.158	37.852	46.509
	10 Fold	0.363	36.254	45.113
LGBM	Train	0.457	32.493	41.85
	Test	0.087	39.734	48.418
	10 Fold	0.325	37.246	46.433
RF	Train	0.423	34.393	43.135
	Test	0.121	38.951	47.513
	10 Fold	0.345	36.986	45.745

Table S8 Comparison between the predictions of the EXP XRD model and new experimental data reported in 2025 that are not included in the training dataset.

Formula	Catalyst Mass Fraction(wt%)	Heating Rate(℃/min)	Experimental Tp(°C)	Predicted Tp(°C)	Ref.
FeMoO@NCN	10	3	270.1	290.8	[11]
		5	278.5	298.6	
		7	289.4	316.4	
		10	295.1	320.5	
NCN	10	3	335.3	346.9	
		5	341.2	352.3	
		7	360.5	380.3	
		10	369.7	389.9	
Co/C	10	5	323.2	313.5	[12]
		8	329.9	321.9	
		10	335.3	327.2	
		12	341.2	335.7	
Ti_2NTx	10	9	374	383.4	[13]
		12	382.3	403.7	
		15	392.1	407.4	
		20	400.1	415.5	
TiO_2-C-Ni	8	5	245	228.2	[14]
		8	258.5	238	
		10	266.8	251	
		13	277.7	275.6	
$Ni(HCOO)_2$-300rpm+TPF	10	5	275	296.8	[15]
$Ni(HCOO)_2$-500rpm	10	5	279	260.5	
$(NiHCOO)_2$-300rpm	10	5	312	337.3	
NF	10	10	335.2	342.5	[16]
NF-200	10	10	325.7	354.2	
NF-300	10	10	319.2	322.3	
NF-400	10	10	327.3	320.5	
NF-500	10	10	356.8	382.7	
VNbC	5	5	235	211.6	[17]
		10	263	243.9	
Ni/NiO@C	10	5	212.8	247.5	[18]
Ni@SC	5	5	307.3	285	[19]
		10	332.7	319.2	
		15	347.3	339.4	
		20	360.5	350.7	
$FeVO_4$/CNT	6	5	307.3	314.9	[20]
		10	332.7	348.5	
		15	338.8	349.7	

		20	342.6	352	
		2	256.9	289.6	
NiO	5	5	277.9	315.1	[21]
		8	291.7	335.6	
		10	302.9	349.8	
		5	351.3	363	
$CuMoO_4$	15	8	359.8	368.3	[22]
		10	367.5	383.8	
	3		307.3	329.7	
	5		294.8	313	
TiO_2/C/Ni		10			[23]
	8		266.8	293.4	
	10		257.6	272.9	
Ni&C	10	5	320.5	344.1	[24]

Table S9 Promising catalyst candidates for MgH_2 dehydrogenation identified by high-throughput prediction.

Formula	Catalyst Mass Fraction(wt%)	Heating Rate(℃/min)	Predicted Tp(°C)
$CaFe_4O_7$	10	2	197.0
$NiZr_2$	10	2	204.4
$Ru_{10}Y_{10}C_{19}$	10	2	206.5
$NaFeF_3$	10	2	215.2
$TbMn_6Ge_6$	10	2	215.3
$Zn_2Co_3TeO_8$	10	5	217.0
$Pr_2O_2SO_4$	10	2	217.3
SnS_2	5	2	218.9
K_2TiF_6	10	2	220.6
$SrZnO_2$	10	2	220.8
$Bi_4(GeO_4)_3$	10	2	222.0
$Nd_2Ti_4O_{11}$	10	2	222.2
$Cu_5V_2O_{10}$	10	2	222.3

Supplementary References

[1] Z. Ding, H. Li, L. Shaw, New insights into the solid-state hydrogen storage of nanostructured LiBH4-MgH2 system, Chemical Engineering Journal 385 (2020) 123856. https://doi.org/10.1016/j.cej.2019.123856.

[2] W. Zhao, Y. Wu, P. Li, L. Jiang, X. Qu, Enhanced hydrogen storage properties of $1.1MgH_2$ –$2LiNH_2$ – $0.1LiBH_4$ system with $LaNi_5$ -based alloy hydrides addition, RSC Adv. 8 (2018) 40647–40654. https://doi.org/10.1039/C8RA07279E.

[3] H. Liu, X. Wang, Y. Liu, Z. Dong, G. Cao, S. Li, M. Yan, Improved hydrogen storage properties of MgH2 by ball milling with AlH3: preparations, de/rehydriding properties, and reaction mechanisms, J. Mater. Chem. A 1 (2013) 12527. https://doi.org/10.1039/c3ta11953j.

[4] C. Duan, M. Wu, Y. Cao, D. Fu, Y. Zhang, Z. Su, Z. Sun, Y. Wu, Novel core–shell structured MgH_2 /AlH_3 @CNT nanocomposites with extremely high dehydriding–rehydriding properties derived from nanoconfinement, J. Mater. Chem. A 9 (2021) 10921–10932. https://doi.org/10.1039/D1TA01938D.

[5] H. Liu, X. Wang, Y. Liu, Z. Dong, S. Li, H. Ge, M. Yan, Microstructures and Hydrogen Desorption Properties of the MgH_2 –AlH_3 Composite with NbF_5 Addition, J. Phys. Chem. C 118 (2014) 18908–18916. https://doi.org/10.1021/jp505064s.

[6] E.M. Dematteis, M. Baricco, Hydrogen Desorption in Mg(BH4)2-Ca(BH4)2 System, Energies 12 (2019) 3230. https://doi.org/10.3390/en12173230.

[7] C. Zuo, C. Zhang, Standardizing differential scanning calorimetry (DSC) thermal decomposition temperatures at various heating rates of an energetic material as a threshold one, Energetic Materials Frontiers (2024) S266664722400054X. https://doi.org/10.1016/j.enmf.2024.06.006.

[8] L. Zhang, X. Zhang, W. Zhang, F. Fang, J. Li, J. Hu, C. Gu, W. Sun, M. Gao, H. Pan, Y. Liu, *In situ* creation of a catalytic multiphase and multiscale surroundings for remarkable hydrogen storage performance of MgH_2, J. Mater. Chem. A 12 (2024) 2423–2434. https://doi.org/10.1039/D3TA06754H.

[9] Q. Yuan, C. Peng, C. Yang, Y. Li, Q. Zhang, Y. Lv, G. Liu, D. Liu, Facilitated hydrogen storage properties of MgH2 by Ni nanoparticles anchored on Mo2C@C nanosheets, International Journal of Hydrogen Energy 85 (2024) 12–19. https://doi.org/10.1016/j.ijhydene.2024.08.292.

[10] T. Tian, X. Wang, K. Shi, F. Liu, J. Wang, X. Zhang, S. Li, W. Chen, Y. Song, J. Li, G. Liu, Boosting hydrogen storage performance of MgH2 by efficient V2O3/C catalyst, Journal of Alloys and Compounds 1010 (2025) 178333. https://doi.org/10.1016/j.jallcom.2024.178333.

[11] J. Zhang, L. Bian, N. Zhang, B. Wang, S. Cao, X. Wang, Y. Zhao, L. Shao, R. Yang, Y. Wang, H. Du, J. Yan, Y. Zhang, X. Lin, J. Zou, Enhancing hydrogen sorption performances of MgH2 by in-situ introduction of molybdate derived FeMo nano catalysts, Journal of Energy Storage 130 (2025) 117418. https://doi.org/10.1016/j.est.2025.117418.

[12] Q. Hou, J. Wang, Y. Zhou, P. Jiang, Y. Li, Z. Ding, Q. Wang, Y. Xiong, Sustainable synthesis of Co/C nanocomposites from biomass for advanced MgH2-Based hydrogen storage: Insights into phase transformation and catalytic mechanisms, Journal of Power Sources 640 (2025) 236713. https://doi.org/10.1016/j.jpowsour.2025.236713.

[13] L. Qin, G. Zhang, T. Liang, L. Sun, F. Xu, Y. Xia, S.P. Verevkin, S.V. Vostrikov, A.L. Maximov, X. Xiao, L. Chen, H. Pan, Oxygen vacancy-rich Ti2NTix MXene as a high-efficiency catalyst for

enhanced hydrogen desorption of MgH2, Journal of Alloys and Compounds 1030 (2025) 180702. https://doi.org/10.1016/j.jallcom.2025.180702.

[14] L. Wang, B. Zhao, J. Liu, J. Yuan, Y. Zhu, B. Liu, Y. Wu, L. Li, Y. Cheng, S.X. Zhou, Effect of Ti-EG-Ni Dual-Metal Organic Crystal-Derived TiO_2 /C/Ni on the Hydrogen Storage Performance of MgH_2, ACS Appl. Mater. Interfaces 17 (2025) 15274–15286. https://doi.org/10.1021/acsami.4c18239.

[15] G. Gao, J. Xie, L. Zhang, C. Lv, C. Li, M. Fan, Z. Yao, Improvement on hydrogen storage performance of MgH_2 by THF-promoted nano-crystallization under low-speed ball milling, Rare Metals 44 (2025) 6366–6374. https://doi.org/10.1007/s12598-025-03330-x.

[16] Y. Chen, X. Li, B. Sun, G. Zhang, S. Ni, C. Li, Effect of NiFe-LDH/FeO(OH) derivatives on the hydrogen storage properties of MgH2, International Journal of Hydrogen Energy 121 (2025) 326–336. https://doi.org/10.1016/j.ijhydene.2025.03.367.

[17] Y. Zheng, L. Zhang, H. Zheng, J. Chen, Z. Chen, Y. Liu, J. Tu, C. Gu, Catalytic effects of multiple heterointerfaces on the hydrogen storage properties of magnesium hydride, Journal of Colloid and Interface Science 685 (2025) 65–72. https://doi.org/10.1016/j.jcis.2025.01.100.

[18] Q. Zhang, Y. Li, F. Sun, X. Lin, Z. Hu, H. Yang, J. Zou, Boosting hydrogen storage performances of MgH2 by using a Ni-MOF derived Ni/NiO@C composite containing Ni/NiO nanoheterojunctions, Chemical Engineering Journal 525 (2025) 169631. https://doi.org/10.1016/j.cej.2025.169631.

[19] Y. Huang, J. Chang, Y. Zhang, X. Zhang, Y. Wang, H. Wu, C. Xu, C. Huang, Z. He, In-situ exploitation of sulfur from petroleum coke in Ni@SC synthesis to Boost the De-/Hydrogenation performance of MgH2: Experimental and DFT study, Fuel 389 (2025) 134568. https://doi.org/10.1016/j.fuel.2025.134568.

[20] R. Zhang, H. Zhang, W. Sun, Z. Yang, X. Lin, W. Guo, Y. Liu, Z. Liang, B. Li, L. Sun, T. Yu, F. Xu, Effect of CNT-referenced FeVO4 as a catalyst on the hydrogen absorption and desorption characteristics of MgH2, Journal of Alloys and Compounds 1029 (2025) 180767. https://doi.org/10.1016/j.jallcom.2025.180767.

[21] T. Tian, F. Liu, X. Zhang, J. Wang, S. Li, W. Chen, Y. Song, J. Li, G. Liu, Effect of porous nanosheet NiO on hydrogen storage performance of MgH2, International Journal of Hydrogen Energy 135 (2025) 2–9. https://doi.org/10.1016/j.ijhydene.2025.05.017.

[22] Y. Pan, X. Yang, X. Lu, J. Su, J. Kong, D. Huo, W. Zhang, Q. Hou, Enhancement catalysis of layered CuMoO4 on hydrogen storage performance of MgH2, Journal of Alloys and Compounds 1017 (2025) 179129. https://doi.org/10.1016/j.jallcom.2025.179129.

[23] L. Wang, B. Zhao, J. Liu, J. Yuan, Y. Zhu, B. Liu, Y. Wu, L. Li, Y. Cheng, S.X. Zhou, Effect of Ti-EG-Ni Dual-Metal Organic Crystal-Derived TiO_2 /C/Ni on the Hydrogen Storage Performance of MgH_2, ACS Appl. Mater. Interfaces 17 (2025) 15274–15286. https://doi.org/10.1021/acsami.4c18239.

[24] S. Zhao, T. Wu, K. Wang, Q. Wang, X. Mei, J. Huang, H. Liu, C. Huang, J. Guo, Synergistic dual-modification regulation of dehydrogenation kinetics of MgH2 by MOF-derived Core-shell Ni&C, International Journal of Hydrogen Energy 127 (2025) 241–251. https://doi.org/10.1016/j.ijhydene.2025.04.125.