Journal of Materials Chemistry A

ROYAL SOCIETY OF CHEMISTRY

PAPER

Cite this: *J. Mater. Chem. A*, 2026, **14**, 7579

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] and Zhengyang Gao[ab]

MgH_2 offers high hydrogen capacity but suffers from an excessively high desorption temperature, which severely restricts its practical application. Developing effective catalysts is essential, yet their structural complexity makes quantitative description and performance prediction notoriously difficult. To overcome the long-standing challenge of catalyst representation, we extracted structural fingerprints directly from standardized X-ray Diffraction (XRD) patterns using dimensionality reduction. These XRD features integrated with catalyst mass fraction and heating rate, were employed to build the machine learning models in the field of MgH_2 catalysis. To predict the dehydrogenation peak temperature (T_p) and its variation (ΔT_p), we compiled a curated dataset from 420 experimental publications, containing over 2000 records. Ensemble regression models were trained separately on experimental and ICDD-based XRD data. Model interpretation using SHAP highlights that XRD-derived features dominate prediction accuracy, while catalyst loading and heating rate show nonlinear but interpretable effects. Based on ICDD-based XRD data, we developed a high-throughput screening workflow coupled with t-Distributed Stochastic Neighbor Embedding visualization, enabling the rapid evaluation of over 100 000 ICDD XRD entries and the identification of 12 promising catalyst candidates (T_p < 230 °C). Based on experimental XRD data, we established a predictive framework for dehydrogenation peak temperatures, capable of integrating user-provided catalyst XRD data. Importantly, we implemented this workflow in XPEAK (an XRD-driven machine learning platform for predicting peak temperatures, http://cat-mh.top), the first platform providing the research community with practical AI-assisted tool for catalyst screening, performance prediction, and design for MgH_2 dehydrogenation. This work demonstrates how XRD fingerprints can resolve the bottleneck of catalyst description and establish a new pathway for data-driven catalyst design.

Received 9th October 2025
Accepted 13th January 2026

DOI: 10.1039/d5ta08221h

rsc.li/materials-a

1. Introduction

MgH_2 has a high theoretical hydrogen capacity of 7.6 wt%,[1–4] while catalysts enhance its kinetics by providing additional active sites, weakening the Mg–H bond, and promoting hydrogen diffusion and surface reactions.[5] In recent years, various strategies have been developed to significantly reduce the dehydrogenation peak temperature (T_p) of MgH_2, thereby enhancing its practical applicability. Zhang *et al.* highlighted in their review that nanostructuring and phase transition engineering can effectively tune the thermodynamic properties of the Mg/MgH_2 system without compromising hydrogen storage capacity.[6] For instance, Liang *et al.* constructed *in situ* Ni/Ni_xB dual-functional catalysts;[7] Li *et al.* introduced single-atom Ni interface engineering;[8] and Ren *et al.* confined MgH_2 within MOF-derived pCNF.[9] Although these strategies reduce T_p by approximately 100 °C, the operating temperature remains above 200 °C, which is still relatively high.

However, traditional catalyst development still relies heavily on empirical trial-and-error, limiting the efficient and systematic exploration of the vast catalyst design space.[10–13] With advances in computational methods and artificial intelligence,[14–19] data-driven models can learn structure–property relationships from existing datasets, enabling high-throughput screening and rapid performance prediction. Such models achieve speed gains of over four orders of magnitude compared to conventional approaches,[20] while maintaining

[a]*Department of Power Engineering, School of Energy, Power and Mechanical Engineering, North China Electric Power University, Baoding, 071003, China. E-mail: yangwj@ncepu.edu.cn*

[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*

accuracy and capturing key composition–structure–property correlations.[21–23] Katarina *et al.* employed a graph neural network approach within the MEGNet framework to achieve high-precision predictions of formation energies.[24] Matthew *et al.* used the Magpie tool to extract elemental and electronic structure features of materials and combined them with gradient boosting tree models to construct a regression model for predicting platform pressure.[25]

Despite the existence of large materials databases such as OQMD,[26] Materials Project, International Centre for Diffraction Data (ICDD), and ICSD, open-source resources dedicated to solid-state hydrogen storage remain scarce. Only a few databases, such as HyDPark, support data-mining studies on hydrogen storage alloys,[27–30] yet no database focuses on catalyst effects.[31] Experimental literature provides a low-cost and high-quality data source. ML models trained on experimental data can offer direct insights into observed behaviors and guide materials design.[32–34]

However, beyond data scarcity, most current machine learning studies on solid-state hydrogen storage materials rely on features that inadequately represent crystal structures and struggle with structurally complex or non-standard materials.[35] Magpie features are based on elemental statistics, lacking information on crystal symmetry and atomic arrangements; one-hot encoding is sparse and contains no structural information; SMILES representations are mainly restricted to organic ligands and are ill-suited for inorganic solids. These limitations hinder them in describing catalyst structures.

To improve the generalization ability and practical applicability of models, feature extraction should be increasingly based on experimental characterization results, as experiments are currently the only means of revealing non-ideal structures and complex behaviors of materials. Among these, X-ray diffraction (XRD) is widely used in material characterization and has been gradually applied to various machine learning tasks in recent years. Oviedo *et al.* developed a physics-informed data-augmentation strategy combined with a fully convolutional neural network, enabling rapid and highly accurate classification of small thin-film X-ray diffraction datasets.[36] Lee *et al.* employed a fully convolutional architecture together with large-scale synthetic data augmentation that accounts for peak perturbations, achieving an accuracy of 90.38% in crystal-system classification on real experimental XRD data.[37] However, Prior studies have demonstrated the effectiveness of XRD data for materials classification and phase identification,[38–44] their applications to domain-specific problems remain limited. Compared to high-dimensional and complex characterization techniques such as electron microscopy,[45] XRD patterns provide more consistent and stable feature representations, while encoding rich information on phase composition, structural characteristics, and certain aspects of processing history. Owing to their concise form yet rich informational content, XRD patterns demonstrate greater applicability and improved modeling efficiency in small-sample learning scenarios.[46–49] Using XRD patterns as machine learning descriptors addresses the limitation in MgH_2-related studies where other approaches fail to characterize catalysts that cannot be represented solely by chemical formulas. This strategy expands the feature space from elemental attributes to experimentally derived phase and structural information. Although XRD-based representations are not new, their systematic application in AI-driven MgH_2 research and catalyst studies remains limited.

Driven by the needs of solid-state hydrogen storage research, this study systematically collected literature and extracted over 2000 differential scanning calorimetry (DSC)-measured T_p and ΔT_p as target variables, along with relevant environmental features. These data were incorporated into our parallel database, *Digital Hydrogen-S*. Experimental XRD (EXP XRD) and ICDD XRD data were used as structural descriptors. After spectral matrix standardization, regression models were developed using ensemble methods.[50,51] In parallel, a convolutional neural network (CNN) regressor was trained for benchmarking, and class-activation-based XRD attribution maps were employed to elucidate the roles of different phases and grain boundaries in catalytic behavior. Subsequently, t-distributed stochastic neighbor embedding (t-SNE) was applied to select 5000 structurally similar samples from over 100 000 candidates, which were rapidly evaluated using the model. To facilitate broader accessibility, an online interface was provided as a convenient resource for researchers.

2. Database and method

2.1 Framework

All data processing and modeling tasks in this study were completed using Python and its third-party libraries. The extraction of experimental literature data and regression modeling can be performed in a standard computer environment. However, for data processing tasks involving large databases such as ICDD, high-performance computing servers were used, along with multiprocessing and memory optimization strategies, to enable efficient parallel processing and significantly improve data processing speed.

The overall workflow of this study is illustrated in Fig. 1, the workflow begins with literature mining and data extraction from both research publications and large-scale databases. For each material, XRD data are extracted as coordinate pairs of diffraction angle and relative intensity, which are then processed into one-dimensional vectors on a standardized angle grid through a normalization procedure. The aligned XRD data are subjected to dimensionality reduction, and combined with target variables and auxiliary features to form a structured dataset. After dataset compilation and partitioning, machine learning models are trained with iterative hyperparameter tuning. For prediction, a test set derived from ICDD entries is screened for similarity and fed into the ICDD-trained model, with radioactive or invalid samples removed from the final results.

2.2 Dataset construction

Data for regression model development were systematically collected from 420 publications on catalyst-modulated MgH_2

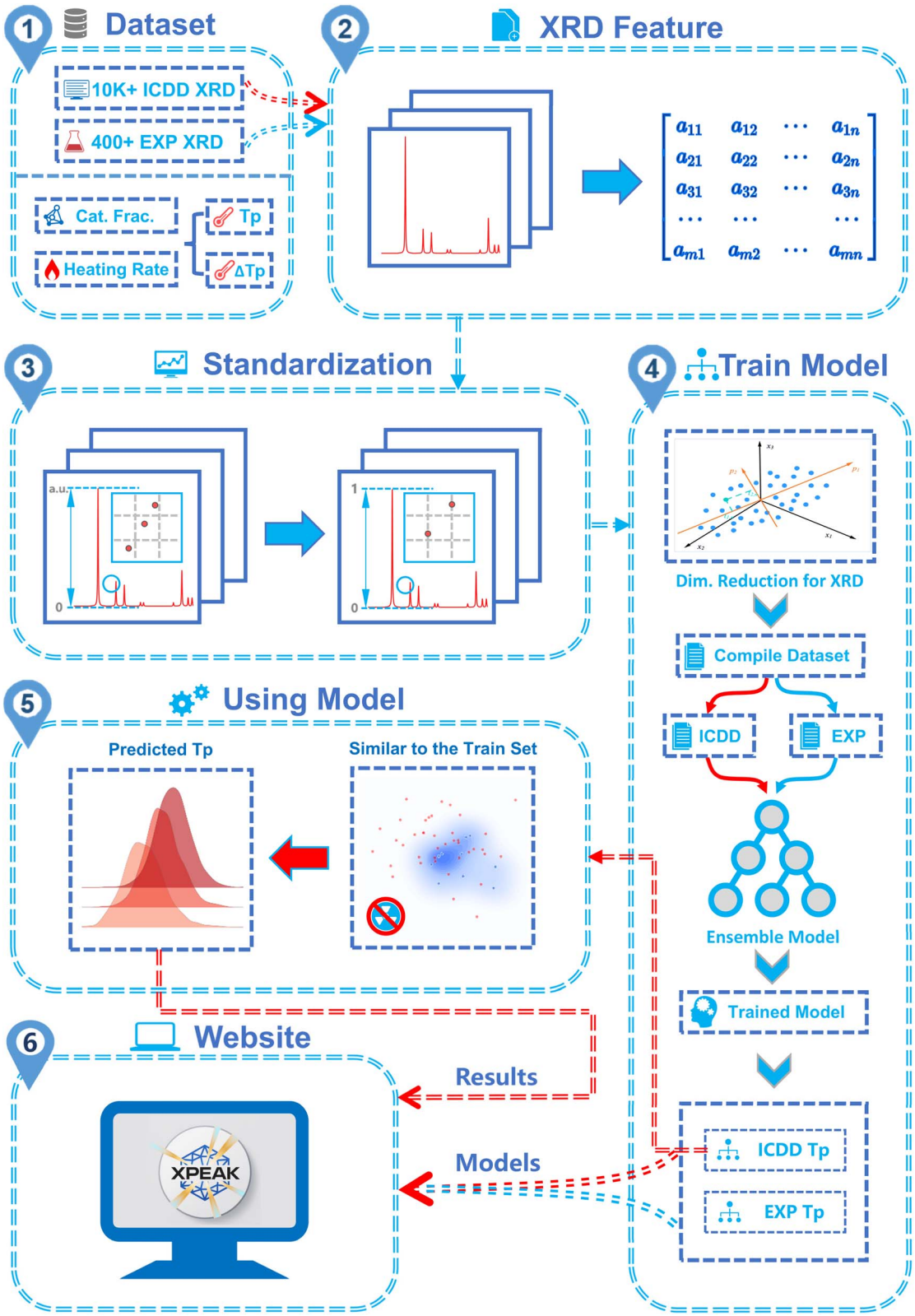

Fig. 1 Schematic diagram of the research workflow, from dataset construction to data processing, model building, and application.

hydrogen storage and release performance spanning the years 2005 to 2024. In total, 2011 DSC-derived samples were compiled, including both T_p and ΔT_p values for 1349 samples, together with information on heating rate and catalyst dosage. Furthermore, over 100 000 standard reference phase XRD patterns were extracted from the ICDD database, and an

additional 393 experimental catalyst patterns were extracted from the paper and used as structural features for model training and prediction. The *Digital Hydrogen-S* platform is a database website developed in our separate work, the procedures for literature retrieval, screening, data digitization, and integration into the *Digital Hydrogen-S* platform are described in the Method S1.

The dataset includes the title, DOI, journal, and publication year of the corresponding literature for each sample to ensure traceability and facilitate data verification. Due to limitations in current multimodal data-recognition technologies, all T_p and XRD data were manually extracted using WebPlotDigitizer.[52] XRD patterns were digitized by color-indexed curve recognition and converted into quasi-continuous coordinate pairs.

2.3 Extracting data from the ICDD database

XRD is a technique that utilizes the interaction between X-rays and crystalline materials to form diffraction peaks.[53] Its fundamental principle is described by Bragg's law:

$$n\lambda = 2d\sin\theta \quad (1)$$

where n is the diffraction order, typically set to $n = 1$, λ is the X-ray wavelength, d is the interplanar spacing, and θ is the incident angle, which is also half of the diffraction angle.

In this study, the experimental diffraction patterns of materials were batch-extracted from the ICDD database using the PDF4+ software. To achieve structured processing, the extracted data were saved in XML files, with the copper target excitation wavelength (Cu Kα_1, λ = 1.54056 Å) consistently selected. Details of the normalization procedures applied to all datasets are presented in Method S2.

2.4 Construction and utilization of the regression model

After initial Standardization, the XRD spectra were resampled to a common grid, resulting in a 4251-dimensional intensity vector. To reduce the dimensionality of the XRD matrix, kernel principal component analysis (kPCA) and non-negative matrix factorization (NMF) were employed.[54] Dimensionality reduction techniques were applied with the objective of preserving 95% of the original information. The maximum number of iterations for the NMF algorithm was set to 2000 to ensure convergence. To prevent data leakage, the training and test sets were strictly separated throughout the modeling process.

Due to the limited scale of domain literature data, which makes it difficult to support the training of deep neural networks. Ensemble models, including XGBoost (XGB), LightGBM (LGBM), and Random Forest (RF), were selected to construct the regression models.[55] All models were trained using a 9 : 1 split between the training and test sets, and 10-fold cross-validation was employed to enhance model robustness and generalization performance.

Using ensemble machine learning algorithms, a black-box model was constructed with the following mapping relationship:

$$T_p = f(W_c, \beta, \mathrm{XRD}) \quad (2)$$

In eqn (2), W_c indicates the catalyst mass fraction, β denotes the heating rate, and XRD corresponds to the reduced feature set derived from the XRD patterns.

To increase the confidence in the prediction results, the t-SNE method was used to select samples in the prediction set that are similar to those in the training set. The remaining samples were input into the best-performing ICDD XRD model for prediction.

3. Results and discussion

3.1 Database analysis

To provide an overview of the current status of dehydrogenation kinetics and catalyst research, we constructed visual statistical charts based on the dataset. As shown in Fig. S1, the manually extracted T_p and XRD data demonstrate acceptable error ranges, thereby providing a reliable basis for subsequent analysis. It is worth noting that during the extraction of T_p, the smoothness of the exothermic peak slightly influences the human judgment of peak positions, while the thickness of XRD spectral lines moderately affects the precision of data extraction by the software. As shown in Fig. 2(a), the KDE (Kernel Density Estimate) curves of dehydrogenation T_p and ΔT_p both exhibit an approximately normal distribution trend. The box plot further indicates that in approximately 50% of the experimental samples, the T_p is concentrated between 275.81 °C and 359.59 °C, while half of the catalysts reduced the dehydrogenation temperature by 33.98 °C to 94.00 °C.

The database includes 695 materials used to catalyze the dehydrogenation of MgH_2. These catalysts are broadly classified into transition metals, alloys, oxides, inorganic salts, nitrides and halides, carbon materials, and carbon-supported types. As shown in Fig. 2(b), current research mainly focuses on transition metals and supported catalysts. The former offer high catalytic activity due to their partially filled d orbitals, while the latter improve stability and dispersion through large surface areas and anchoring sites, reducing agglomeration and enhancing kinetics *via* confinement effects. As shown in Fig. 2(c) and S2, Ni, Fe, Ti, and carbon-based supports are the most commonly used catalysts,[56–58] Ti-based and V-based catalysts exhibit low median T_p values and narrow, positively shifted ΔT_p distributions,[59] consistent with their good mutual solubility with Mg, which facilitates the formation of Mg–Ti/V solid solutions or Mg–Ti–H phases at surfaces or grain boundaries.[60] These intrinsic effects are relatively insensitive to microstructural variations, leading to concentrated performance outcomes. Fe-based,Co-based catalysts show much broader ΔT_p ranges, as their activity depends on H_2 dissociation and spillover on surface nanoparticles, which are highly sensitive to particle size, dispersion, and surface chemistry.[61,62] Carbon-supported catalysts display the most favorable overall characteristics, with T_p concentrated in the low-temperature region and ΔT_p distributions that are narrow with high medians, reflecting stable enhancements from dispersion and interfacial synergy. Despite extensive studies on transition metals, their

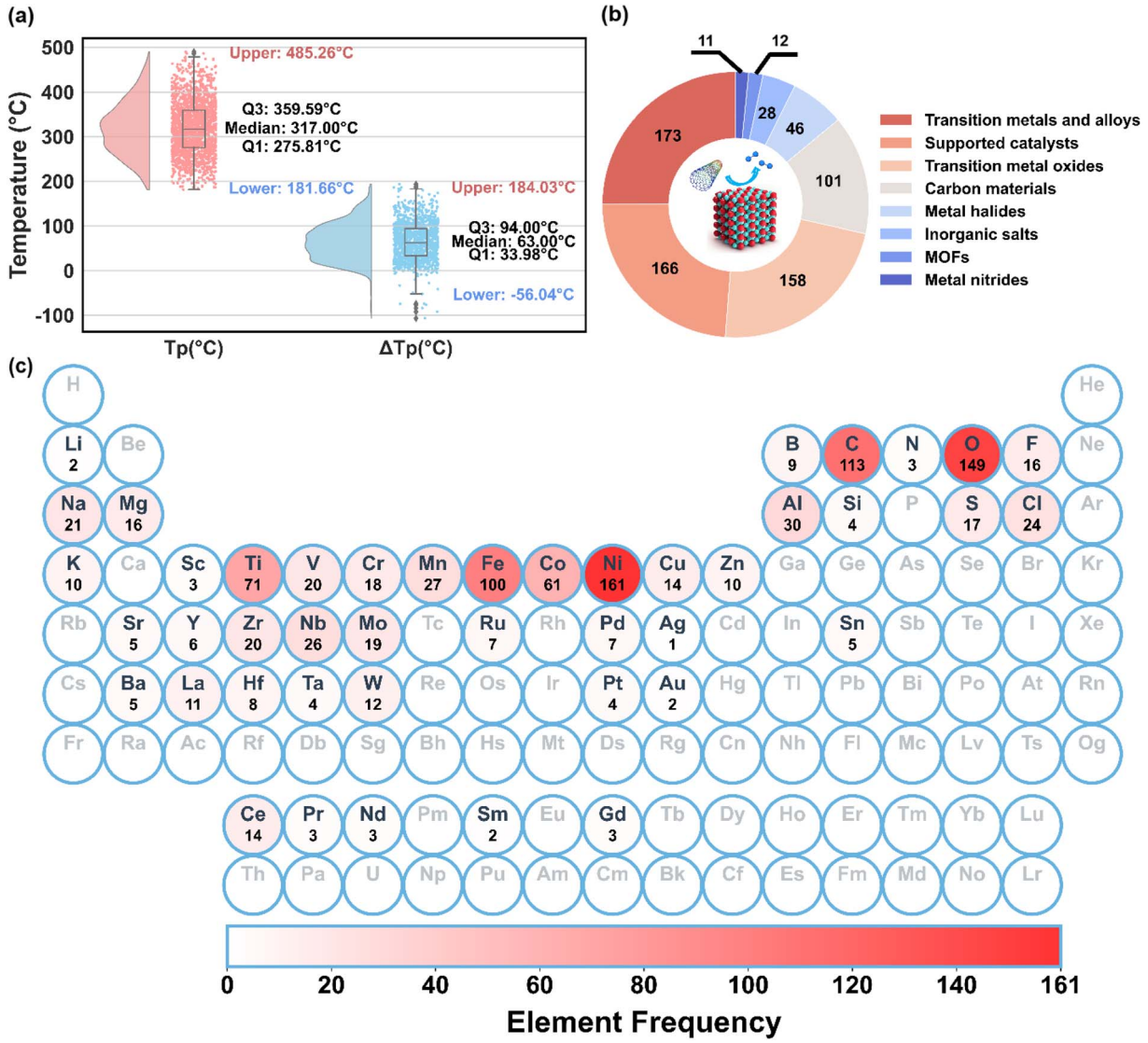

Fig. 2 Analysis of the database extracted from the literature. (a) Cloud plot showing the dehydrogenation T_p (red) and ΔT_p (blue) of MgH_2 after catalyst addition, illustrating the distribution of the two types of data. (b) Analysis of the material research in the database, with the numbers in the pie chart representing the frequency of occurrence of each type of catalyst. (c) Analysis of the elements contained in the catalysts studied in the database, with the numbers under the element symbols indicating the frequency of occurrence for each element.

highly scattered performance distributions indicate significant uncertainty in composition tuning for alloy-type catalyst systems.[63–66]

As shown in Fig. S3, joint dimensionality reduction and clustering analysis of selected catalyst XRD patterns from the database reveals a certain similarity between the clusters generated by unsupervised learning and manually defined material categories. This suggests that XRD data inherently contain sufficient structural information to distinguish intrinsic differences among materials, even in data-scarce scenarios. Meanwhile, the relatively dispersed distribution of supported catalysts indicates potential limitations of current machine learning models in effectively capturing and interpreting the complex features of supported catalyst systems.

3.2 Preprocessing results

The data structure of the XRD patterns still does not fully meet the consistency requirements for input features in machine learning models. In particular, because XRD spectra are continuous variables, the sampling points vary across different data sources. Therefore, interpolation algorithms were applied to re-sample the spectra at equal intervals.

To optimize the interpolation strategy for XRD data, this study evaluated combinations of linear, quadratic, and cubic interpolation methods with varying step sizes on the training dataset, using the global MAE between the interpolated and original spectra as the evaluation metric. As shown in Fig. 3(a) and Table S1, all three interpolation methods applied to standardized ICDD XRD patterns yielded near-zero MAE when the step size was below 0.02°. As shown in Fig. 3(d), for the

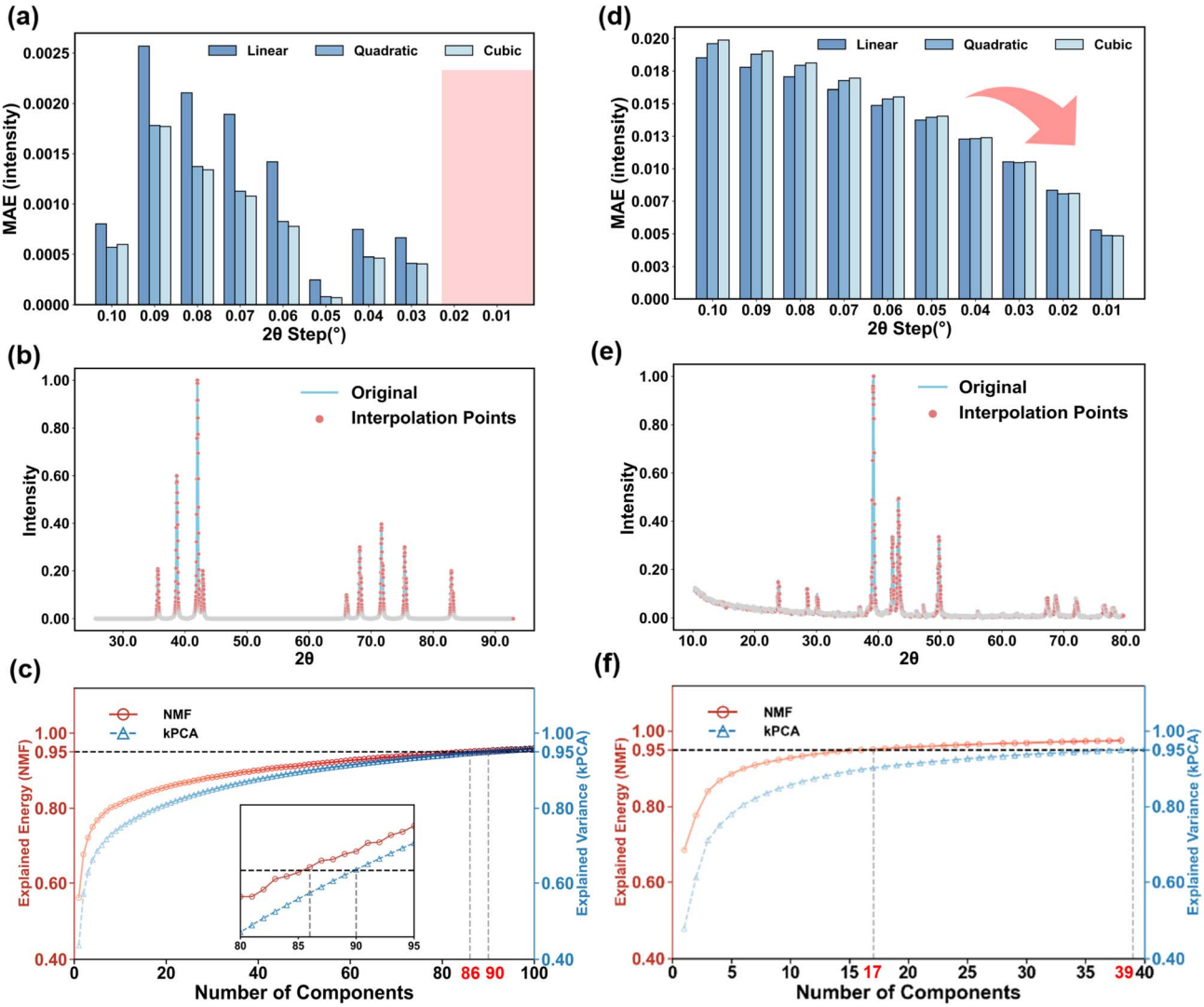

Fig. 3 Results and analysis of XRD normalization and dimensionality reduction. Subfigures (a–c) correspond to ICDD XRD data, while (d–f) pertain to EXP XRD data. Subfigures in the same row represent the same type of analysis: (a and d) Mean reconstruction error of interpolated relative intensities using linear, quadratic, and cubic interpolation methods at varying step sizes; (b and e) visualization of interpolation results for randomly selected samples, where the original spectrum is shown as a light blue line and the interpolated data points are plotted in light red for clearer comparison; (c and f) trends of explained energy for NMF (red) and explained variance for kPCA (blue) as a function of the number of components.

normalization of EXP XRD patterns, the MAE decreased consistently with smaller step sizes across all methods. As shown in Fig. S4, the XRD signals of the training samples are predominantly distributed within $2\theta = 5°–90°$, consistent with established knowledge.

To further validate the structural fidelity of interpolated patterns, Fig. 3(b and e) visualizes linear interpolation results for a randomly selected sample after normalization. The interpolated spectrum retains the main peak positions, shapes, and relative intensities of the original, with no observable peak shifting or artificial peak generation. For consistency, the prediction set of ICDD XRD was processed using the same interpolation method and step size, and representative examples are shown in Fig. S5. As shown in Fig. S6, a KNN imputation combined with a weighted loss function was tested to fill missing boundary points; although it improved local smoothness, it introduced spurious peaks and error accumulation, thereby reducing the physical interpretability of the data.

To mitigate the curse of dimensionality, we performed dimensionality reduction on the XRD feature matrices before model training.[67] As shown in Fig. 5(c and f), NMF reduced the dimensionality of the ICDD XRD data to 86 and that of the EXP XRD data to 17. In comparison, kPCA reduced the dimensionality to 90 for ICDD XRD and 39 for EXP XRD.

As shown in Table 1, dimensionality reduction markedly improves the statistical robustness of the models, with the sample-to-feature ratio (N/P) increasing from below 0.32 in the raw ICDD and EXP XRD data to values exceeding 10 for all NMF and kPCA configurations. This substantial enhancement in N/P

Table 1 Sample-to-feature ratio (N/P ratio) analysis before and after dimensionality reduction

		ICDD		EXP	
Method	Dimensionality processing	T_p (°C)	ΔT_p (°C)	T_p (°C)	ΔT_p (°C)
NMF	Before dimensionality reduction	0.32	0.22	0.17	0.12
	After dimensionality reduction	15.38	10.82	71.21	50.16
KPCA	Before dimensionality reduction	0.32	0.22	0.17	0.12
	After dimensionality reduction	14.71	10.35	33	23.24

effectively mitigates overfitting risk and provides a more stable foundation for ensemble regression modeling.

As shown in Fig. 4 and S7–S11, the low Pearson correlation coefficients between features and the absence of well-fitted regression lines in the scatter plots indicate minimal linear dependencies among features. The environmental variables are largely independent of the structural features, effectively mitigating multicollinearity issues. As shown in Fig. 4(a and b) and S9, aside from a weak linear correlation between heating rate and T_p, no clear linear relationships are observed between T_p and other features. Fig. S10 and S11 further indicate that ΔT_p shows no linear correlation with any feature. Therefore, tree-based regression models are well-suited for this task, as they

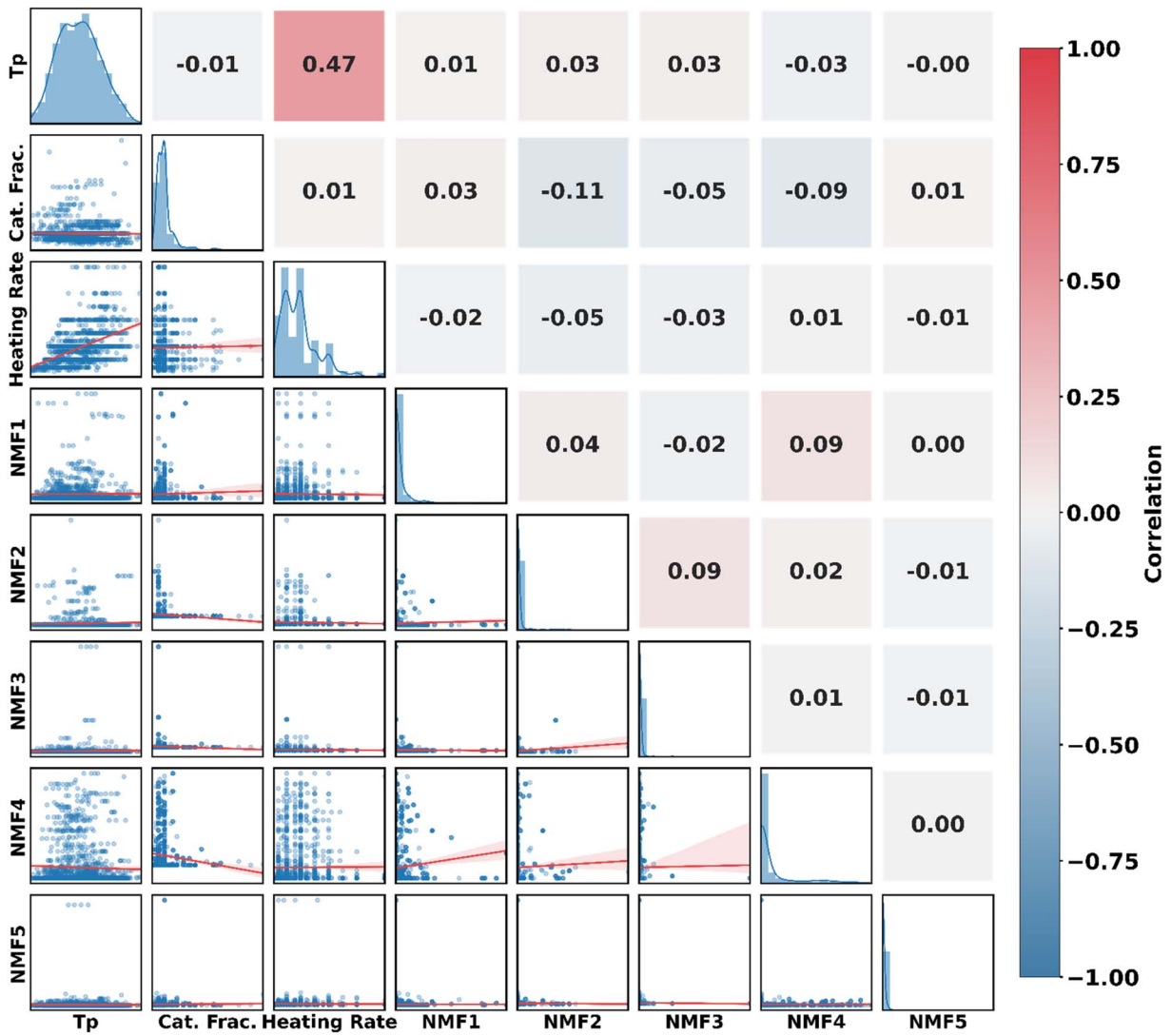

Fig. 4 Relationships among local NMF features, environmental variables, and the T_p. Correspond to ICDD XRD data. Variables are ordered as follows from top to bottom on the left and left to right along the bottom: T_p, Cat. Frac., Heating Rate, NMF1, NMF2, NMF3, NMF4, NMF5. Each panel includes marginal 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.

naturally capture feature interactions and uncover nonlinear relationships between the target variables and predictors.

3.3 Machine learning model construction and analysis

To systematically evaluate the performance of different machine learning models in predicting T_p and ΔT_p, this study constructed six models for each target variable using two dimensionality reduction strategies and three ensemble regression algorithms. In addition, a 1D-CNN model was trained for direct comparison with the ensemble models.

As shown in Table S3, most ensemble models trained using ICDD XRD features performed well on the training set for the T_p prediction task, but their generalization ability varied on the test set and in cross-validation. Considering the metrics in Table S3 and the visualizations in Fig. 5(a and b), the NMF + LGBM model demonstrated the best overall performance, achieving a test $R^2 = 0.848$, and MAE = 12.74 °C. It also maintained stable results in ten-fold cross-validation, with an average $R^2 = 0.822$. The residual distribution shown in Fig. 5(b) further confirms that the errors are relatively small and well-concentrated, with only a few outliers and a slight left skew. As shown in Table S3 and Fig. S13(a1 and a2), the ΔT_p prediction models trained on ICDD XRD data exhibit slightly weaker generalization performance.

Based on the metrics presented in Table S4, ensemble models trained on EXP XRD generally exhibit superior performance. As shown in Table S4 and Fig. 5(c and d), the kPCA + XGB model achieves the best performance in T_p prediction, with an $R^2 = 0.88$ and MAE = 9.94 °C on the test set. It also reaches $R^2 = 0.88$ in ten-fold cross-validation. The residuals show a relatively small standard deviation and a skewness of 0.80, indicating minimal systematic bias. As shown in Table S4 and Fig. S15(e1 and e2), the kPCA + LGBM model performs best in ΔT_p prediction, with a test $R^2 = 0.885$, RMSE = 11.80 °C, and MAE = 7.74 °C. Both ΔT_p and T_p prediction tasks demonstrate a higher level of generalization compared to models based on ICDD XRD.

Compared to ensemble models trained on ICDD XRD data (Table S3 and Fig. 5(a and b), S12–S13), models trained on EXP XRD data (Table S4 and Fig. 5(c and d), S14–S15) demonstrate superior predictive performance despite using fewer samples. This advantage stems from the more comprehensive retention of the intrinsic three-dimensional features of the catalyst in the experimental data: surface geometry and electronic environment capturing lattice stress, strain, solid solution formation, and crystal plane exposure; bulk structure involving grain size, defects, phase transformations, and amorphous/crystalline ratio; and the real material state reflecting microstructural changes induced by preparation and treatment. However, the lack of large, publicly available experimental databases still limits their broader application in high-throughput scenarios.

Ding *et al.* reported model performance test set $R^2 = 0.834$ on the test set. Our model achieves slightly higher predictive accuracy.[68] We argue that the effectiveness of XRD patterns as

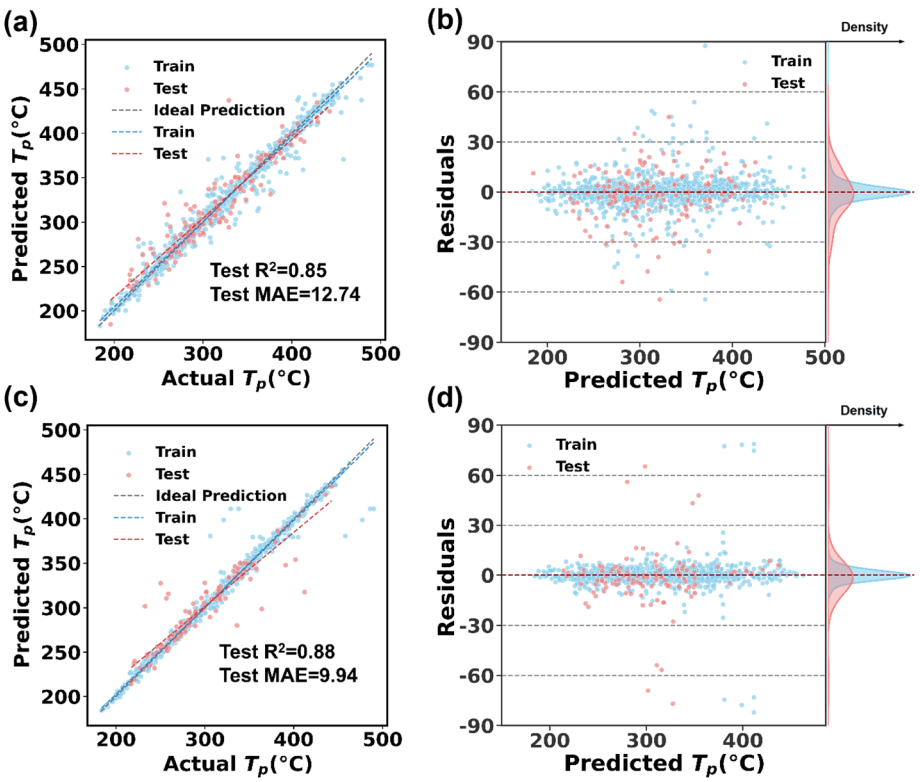

Fig. 5 Performance of the optimal ensemble models for predicting T_p based on ICDD XRD and EXP XRD data. (a) and (b) show the scatter plot and residual distribution with KDE for the model trained with NMF + LGBM to predict T_p with ICDD XRD input data. (c) and (d) show the scatter plot and residual distribution with KDE for the model trained with kPCA + XGB to predict T_p with EXP XRD input data.

input features lies in their ability to encode not only crystal structure and phase information, but also to directly reflect parameters such as lattice constants, phase composition, grain size, crystalline defects, and microstrain through the position, intensity, and shape of diffraction peaks. While idealized structural descriptors or purely atomic graph representations excel at capturing periodic crystal structures and are suitable for predicting intrinsic properties of ideal crystals such as band gaps, XRD inherently incorporates parameters closely related to catalytic performance, such as grain size and lattice defects.[69] This makes XRD features more appropriate for predicting experimentally relevant macroscopic properties like catalytic activity. In addition, to benchmark against conventional feature engineering, we extracted Materials Project descriptors *via* the MP API. As summarized in Table S5, the resulting models reached only ~0.7 R^2 on the training set, substantially lower than those trained on XRD-derived features.

As shown in Table S6, the CNN regression models achieve train set R^2 = ~0.8 and test set R^2 = 0.7–0.8, with generally higher MAE, performing inferior to ensemble tree models. This performance gap arises because neural networks require substantially larger datasets for stable parameter optimization, whereas tree-based models remain effective even under limited sample sizes.

As shown in Table S7, we performed an ablation study by removing all XRD features and retraining the models. The substantial drop in predictive accuracy confirms the critical contribution of XRD-derived information to model performance.

3.4 Feature analysis

Feature importance analysis serves as a key step in understanding the driving factors behind model predictions. By uncovering trends in these latent mechanisms, it provides insights that can inform and guide experimental design. Leveraging a unified theoretical framework, SHAP quantifies the marginal contribution of each feature while accounting for feature interactions; Class activation maps derived from CNN enable visualization of the critical input regions that the model attends to during prediction.

As shown in Fig. 6(a and b), S16, S18(a1–f1), and S22(a1–f1), SHAP analyses indicate that lower heating rates cause the blue data clusters to shift leftward, reducing the predicted values, whereas higher rates drive the red clusters rightward, increasing the predictions. When the heating rate is below 10 °C min^{-1}, the SHAP values increase nearly linearly with the rate, consistent with the behavior described by the Kissinger equation.[70–74] Once the rate exceeds 10 °C min^{-1}, the slope decreases markedly, indicating that heat-transfer delays begin to significantly influence the reaction.[75–78] The model results align with experimental observations, and the heating rate emerges as a highly influential factor in T_p prediction,

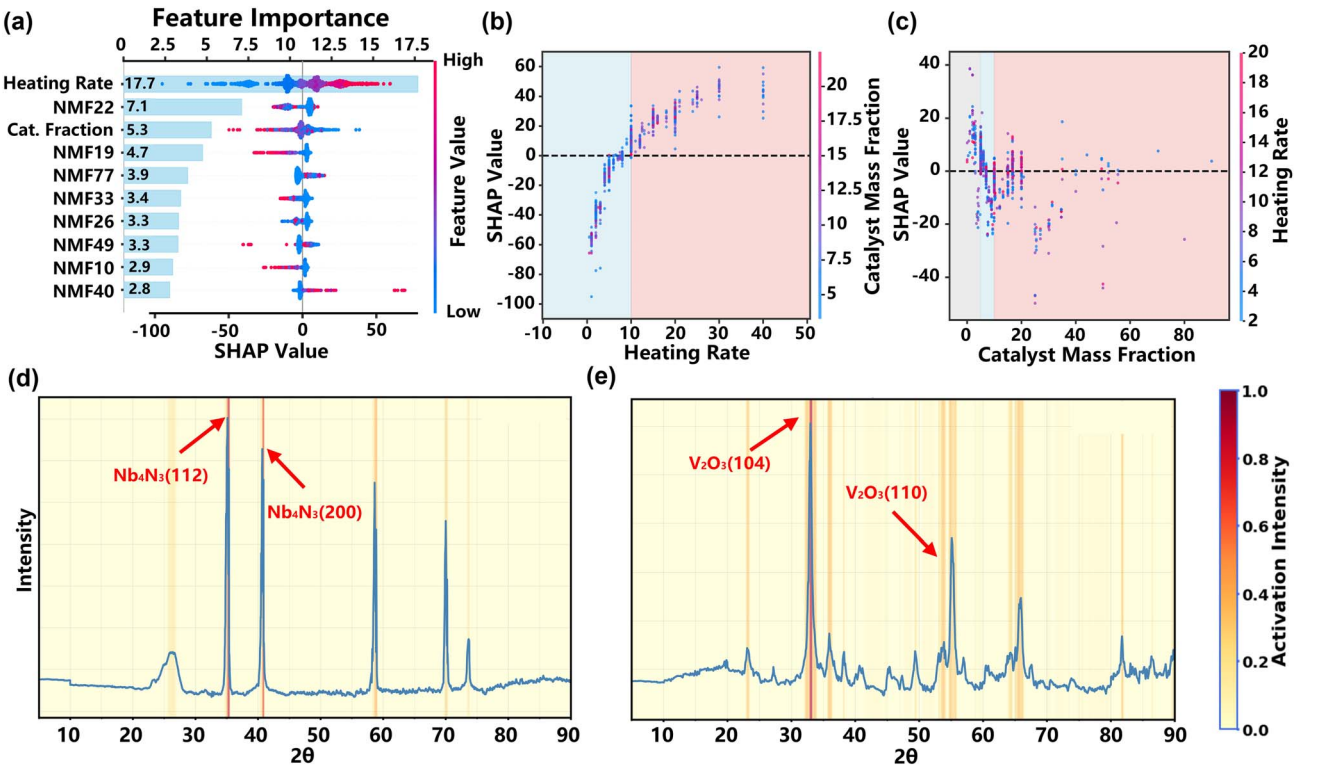

Fig. 6 SHAP-based feature interpretation of the ensemble models: (a) top 10 features ranked by global importance, where bar plots indicate global importance and the adjacent scatter plots show the distribution of SHAP values across samples. (b) Univariate SHAP dependence plot showing how SHAP values vary with heating rate; point colors represent catalyst mass fraction. (c) Univariate SHAP dependence plot showing how SHAP values vary with catalyst mass fraction; point colors represent heating rate. CNN-based class activation map (CAM) interpretation of XRD fingerprint features: (d) Nb_4N_3; (e) V_2O_3.

highlighting the need for more systematic control and standardization in experimental measurements.

As shown in Fig. 6(c), S18(a2–f2), S19(a2–f2), S22(a2–f2), and S23(a2–f2), the overall SHAP trends indicate that low catalyst loadings produce only limited reductions in T_p, likely due to insufficient catalyst dispersion or incomplete interface formation.[79–82] As the loading increases, the SHAP values remain negative but begin to rise, suggesting diminishing catalytic returns, which may result from catalyst agglomeration, excessive surface coverage, or obstructed diffusion pathways.[83–88]

Inspired by Szymanski *et al.*,[40] we have employed class activation maps derived from the CNN to propose exploratory mechanistic interpretations. As shown in Fig. 6(d), the model identifies the (200) and (112) facets of the Nb_4N_3 phase as the structural descriptors governing the catalytic dehydrogenation performance of MgH_2. These specific facets provide optimal interfacial atomic arrangements that support the multivalent Nb^{3+}/Nb^{5+} "electronic pump" mechanism, thereby facilitating efficient electron transfer. Their structurally stable crystal configuration may further ensures durable catalytic activity, consistent with the experimental findings reported by Zhang *et al.*[89] As shown in Fig. 6(e), the diffraction angles corresponding to high-index facets such as (104) and (110) exhibit pronounced activation intensities. In addition, the broadened peak regions display noticeable activation signals, consistent with the 8 nm V_2O_3 nanoparticles and their high specific surface area of 209 $m^2\ g^{-1}$ observed by Wang *et al.*[90] *Via* TEM. This agreement suggests that the machine-learning model may captures the nanoscale effects associated with increased active-site density and optimized surface defects.

Across multiple class-activation maps, the XRD peaks with the highest diffraction intensity consistently align with the most prominent activation regions, implying that the dominant crystal facets of the catalyst play a central role in governing the reaction. These highly ordered surfaces provide stable and efficient pathways for charge transfer and define the primary

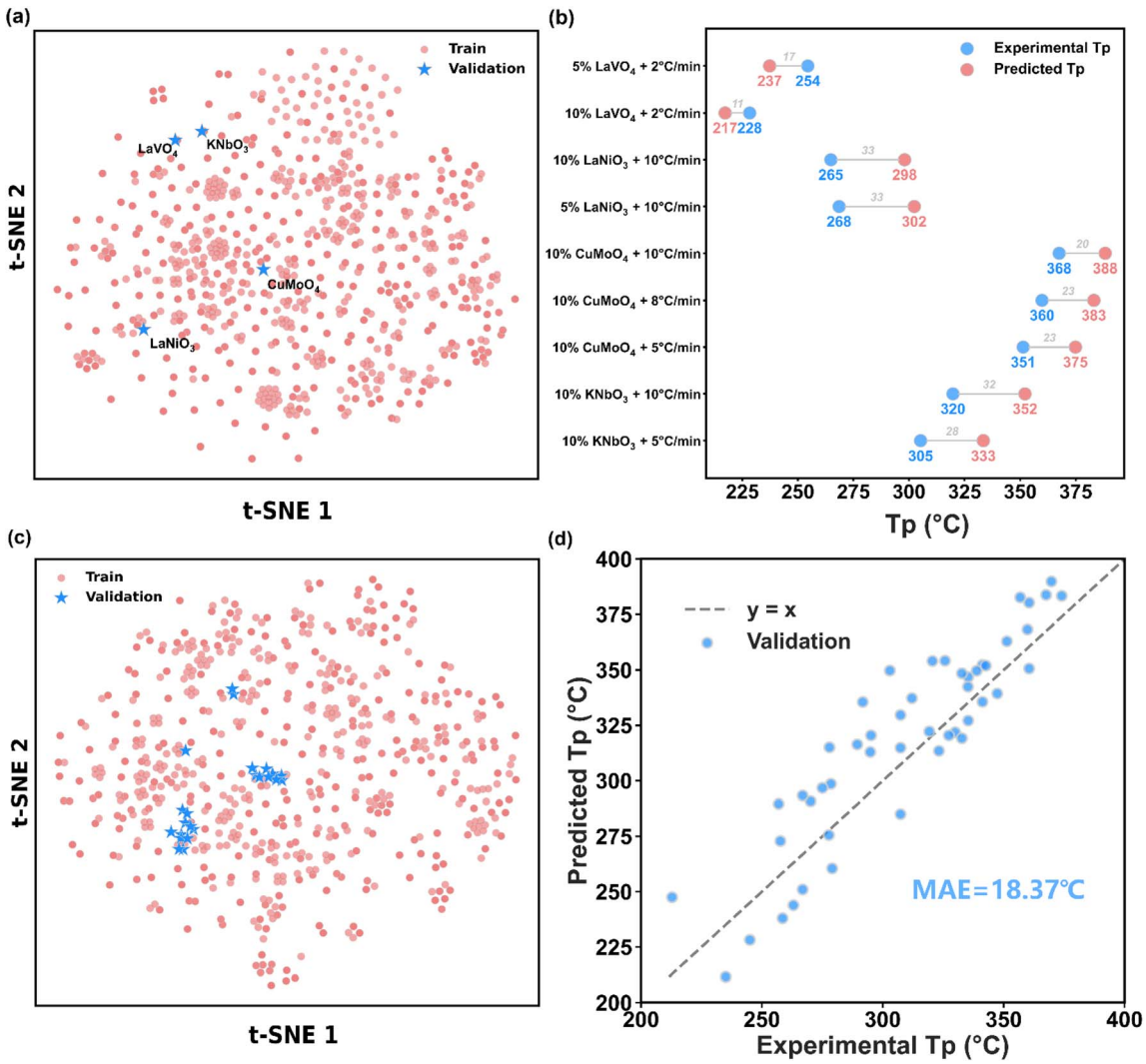

Fig. 7 t-SNE sample similarity analysis and validation of prediction results. (a) Distribution analysis of the shared samples between the newly reported 2025 experiments and the prediction set, alongside the training set samples. (b) Comparison of high-throughput T_p predictions from the optimal ICDD XRD-based model with new experimental results. (c) Distribution analysis of newly tested samples *versus* training set samples. (d) Validation of EXP XRD model predictions against new experimental truths.

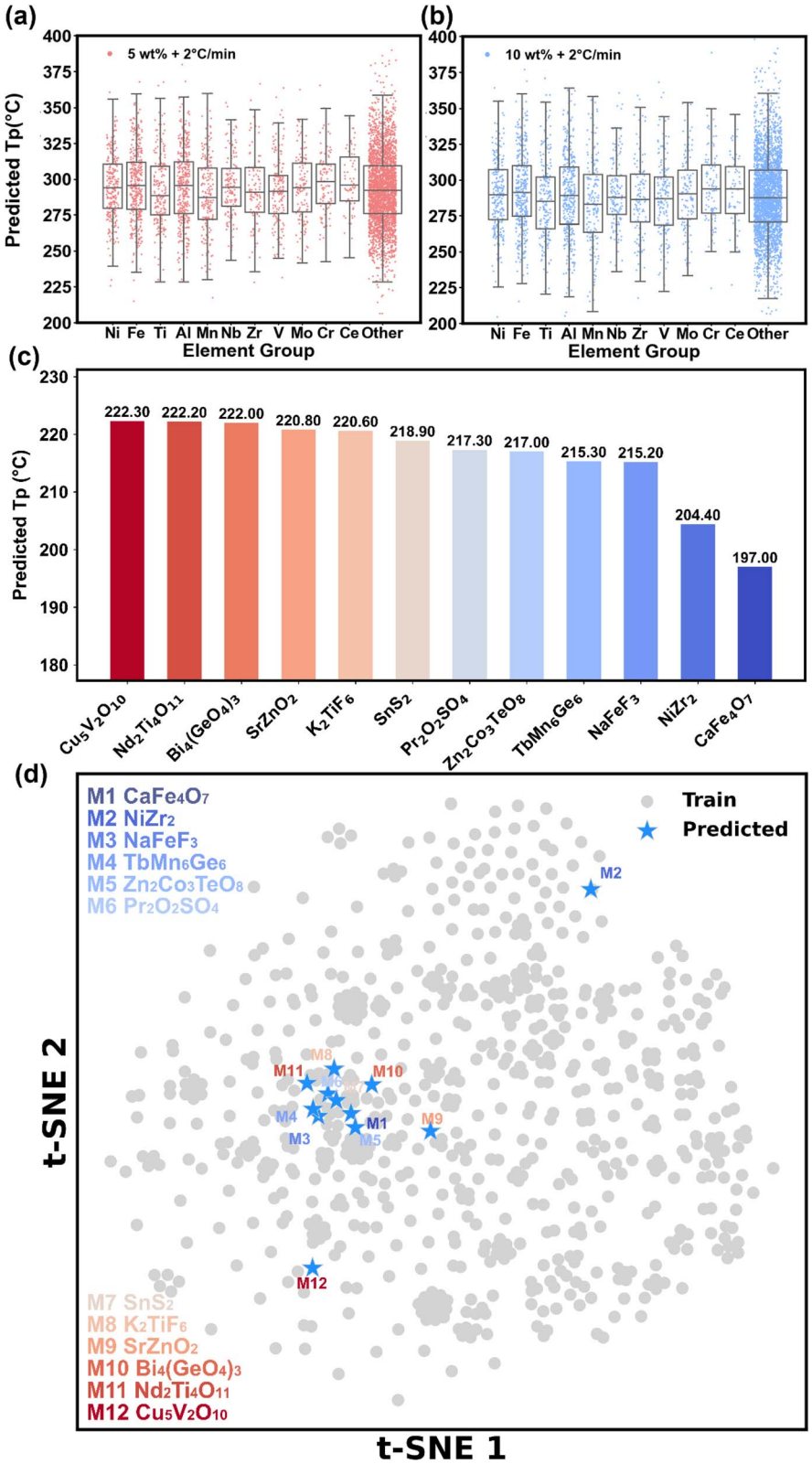

Fig. 8 Predicted performance of 5000 catalyst materials heated at 2°C min^{-1} under (a) 5 wt% and (b) 10 wt% loadings. *Y*-axis elements are common catalyst components; box plots with overlaid scatter points represent materials containing each element. (c) Dehydrogenation temperatures of the promising catalysts identified through high-throughput screening at a heating rate of 2 °C min^{-1} (bar color is for visual distinction only). (d) t-SNE visualization of the distribution of promising catalysts in the prediction set compared with those in the training set.

low-energy channels for Mg–H bond activation. Meanwhile, the secondary activation signals associated with broadened diffraction peaks indicate that nanoscale effects and crystal defects are likewise essential. Defect structures and low-coordination sites can increase the density of local active sites, introduce strain fields, and modulate surface electronic states, thereby enhancing adsorption and bond rearrangement processes and further lowering local reaction barriers. Collectively, the principal crystal facets may establish the fundamental reaction pathways, whereas controllable defects and nanoscale features contribute additional local strengthening effects. These complementary mechanisms operate at different structural levels to jointly promote the dehydrogenation of MgH_2.

3.5 Model utilization

To enable the application of the EXP XRD model, we simulated experimental factors such as grain refinement, preferred orientation, and thermal diffuse scattering on the ICDD XRD spectra. However, as shown in Fig. S24(a), as described in Method S4, the simulated ICDD samples and the experimental samples extracted from literature form two clearly separated clusters, with one being significantly larger than the other. This separation indicates a systematic difference in data patterns between the two sources. The result suggests that current simulation methods in this field are still inadequate for making the prediction set compatible with the EXP XRD model. In contrast, when both the training and prediction sets are

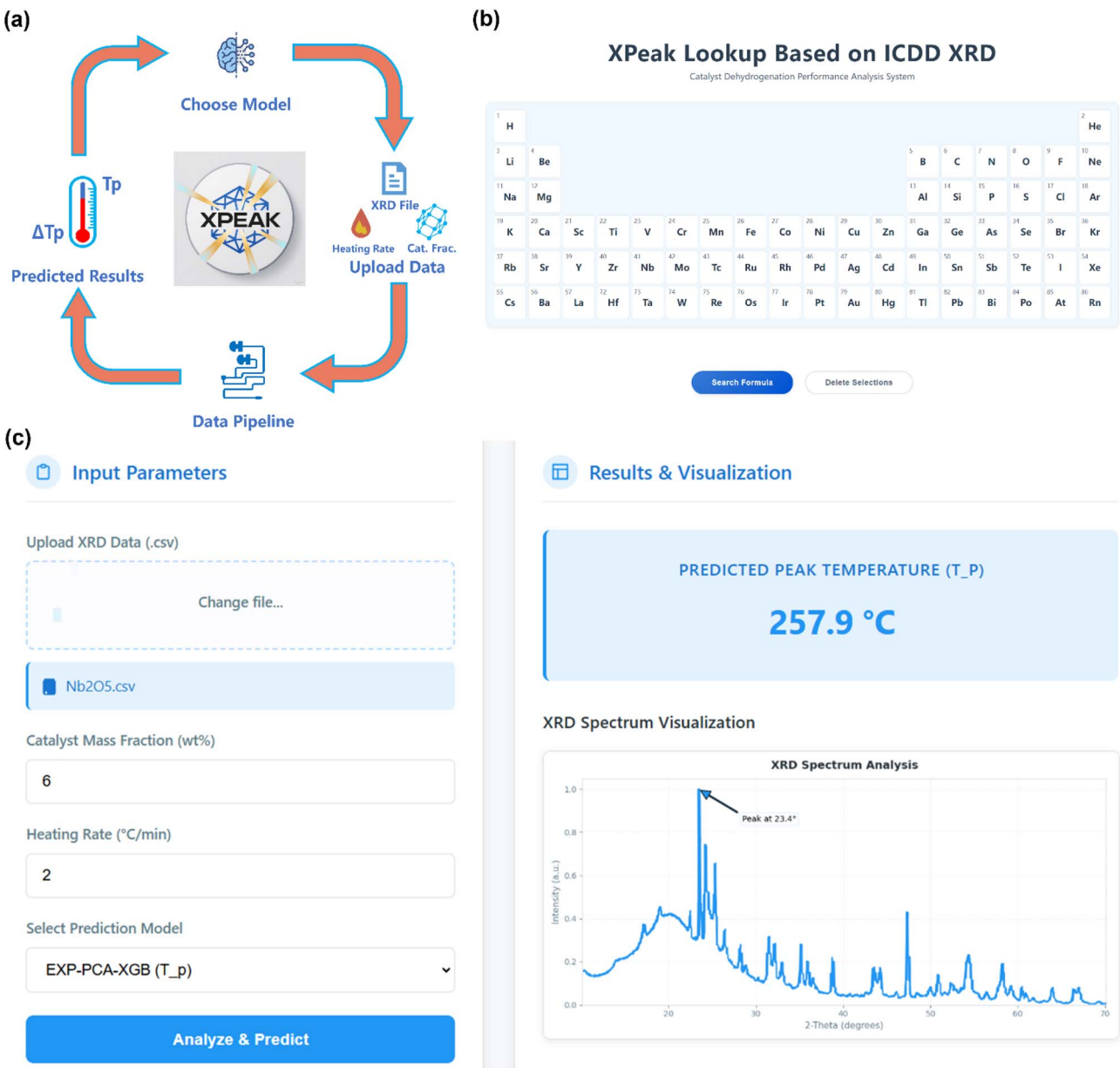

Fig. 9 Schematic diagram of the XPEAK model website. (a) Workflow diagram. (b) High-throughput prediction results query website. (c) Machine learning model website.

constructed using ICDD-derived XRD data, their distributions, as illustrated in Fig. S24(b), tend to be more similar.

t-SNE was employed to identify samples in the high-dimensional space that exhibit structural similarity to the training set. A total of 5000 candidate materials with XRD patterns closely resembling those of the training samples were selected. Model inference was conducted on a 128-core workstation, and all predictions and output processing were completed within 3 minutes.

To provide a more intuitive validation of the high-throughput prediction results, we compared the predictions with data reported in newly published literature from 2025. As shown in Fig. 7(a and b), four catalysts including $KNbO_3$,[91] $CuMoO_4$,[92] $LaNiO_3$ (ref. 93) and $LaVO_4$ (ref. 94) were identified in both the newly published literature and the prediction set. In Fig. 7(a and c), these catalysts exhibit nearly the same distribution as the training set, while in Fig. 7(b), their prediction errors range from 11 to 33 °C. Fig. 7(d) indicates that the model demonstrates acceptable performance on the newly reported literature data, with detailed comparative data provided in Table S8.

As shown in Fig. 8(a and b) and S25, the Ti and Mn groups exhibit significantly lower medians and interquartile ranges compared to other element groups. To further clarify the promising catalysts identified through high-throughput screening, Fig. 8(c) shows their corresponding T_p values at 2 °C min^{-1}, with detailed experimental conditions provided in Table S9. As shown in Fig. 8(d), these promising catalysts exhibit strong similarity in feature-space distribution to the training set, suggesting that their prediction errors are likely to be small.

To facilitate user accessibility, as illustrated in Fig. 9, we designated the optimal combination of prediction models as XPEAK, which has been deployed on the Cat-MH platform and is accessible at http://cat-mh.top. Specifically, Fig. 9(a) presents the input data processing pipeline, developed in strict accordance with the preprocessing procedures described earlier. Fig. 9(b) illustrates the element-indexed query website, which allows readers to conveniently explore the outcomes of our high-throughput predictions, this functional module can be directly accessed at http://cat-mh.top/lookup. Fig. 9(c) shows the T_p and ΔT_p prediction model platform, including a representative case study of applying the EXP XRD model,[95] this module is available at http://cat-mh.top/predict.

4. Conclusion

In this study, we developed the first machine learning model for predicting the T_p and ΔT_p in MgH_2-based systems. This model was trained on experimental data compiled from over 400 published studies, addressing a critical gap in the prediction of key performance metrics within the solid-state hydrogen storage field. By integrating catalyst dosage, heating rate, and XRD structural features, particularly those reduced *via* NMF and kPCA, we demonstrated the feasibility and accuracy of using tree-based models for performance prediction. SHAP-based analysis and CNN class activation inference revealed that XRD-derived structural components play a dominant role. Furthermore, a high-throughput screening workflow was established using ICDD XRD data, enabling efficient performance prediction for 5000 candidate materials. An online platform (http://cat-mh.top) was developed based on the optimal predictive model to facilitate community-wide catalyst sharing.

This work not only provides a structured and experimentally grounded database for MgH_2 catalyst design, but also expands the application of XRD fingerprints in data-driven materials and catalysts research. Based on the intrinsic properties of catalysts, we consider catalyst-related descriptors to be an effective and general-purpose feature in materials machine learning. Since crystallographic fingerprints encode information on phase composition, crystallinity, microstrain, defects, and grain size, XRD-based models can be applied to predict hydrogen sorption kinetics, activation energy, cycling stability, and catalyst-induced structural evolution. Existing regression models in the solid-state hydrogen storage field remain fundamentally limited by insufficient training samples and challenges in data collection. Machine learning in this domain is still emerging. Building upon this work, future research could integrate more comprehensive datasets, develop models better suited for small-sample scenarios, incorporate additional experimental characterization modalities as features, and construct frameworks capable of interpreting experimental data. This will further deepen the application of CAM-assisted XRD in revealing mechanisms, thereby enhancing model interpretability and enabling interpretable optimization of materials in downstream applications.

Author contributions

Weijie Yang: conceptualization, resources, methodology, investigation, formal analysis, writing – review & editing. Zhuoran Xu: conceptualization, data curation, software, formal analysis, visualization, writing – original draft. Jiongyang Li: software, validation. Leping Zhang: software, visualization. Man Shu: investigation, resources. Jianghao Cai: investigation, data Curation. Tongao Yao: resources, project administration. Xiaotian Tang: validation, visualization. Yuxuan Liu: funding acquisition, supervision. Shiwen Luo: funding acquisition, supervision. Xuqiang Shao: supervision. Wenjing Wang: supervision. Zhengyang Gao: resources.

Conflicts of interest

There are no conflicts to declare.

Data availability

All data can be obtained from the corresponding author upon reasonable request. The experimental data extracted from published literature are available through the *Digital Hydrogen-S* Platform (http://digital-hydrogen.com/storage/index-white.html). The XRD data obtained from the International Centre for Diffraction Data (ICDD) are subject to data-use agreements and can be accessed upon formal request to the ICDD. The code and the dimensionality-reduced dataset for

training and testing used in this study can be accessed at https://github.com/Weijie-Yang/XPEAK.

Supplementary information (SI) is available. See DOI: https://doi.org/10.1039/d5ta08221h.

Acknowledgements

This work was supported by Beijing Natural Science Foundation (2262076), Natural Science Foundation of Hebei (E2025502039), Fundamental Research Fund for the Central Universities (2025JC008 and 2025MS131).

References

1 A. P. Zhao, S. Li, D. Xie, Y. Wang, Z. Li, P. J.-H. Hu and Q. Zhang, Hydrogen as the nexus of future sustainable transport and energy systems, *Nat. Rev. Electr. Eng.*, 2025, **2**, 447–466, DOI: 10.1038/s44287-025-00178-2.

2 H. Wang, J. Li, X. Wei, Y. Zheng, S. Yang, Y. Lu, Z. Ding, Q. Luo, Q. Li and F. Pan, Thermodynamic and Kinetic Regulation for Mg-Based Hydrogen Storage Materials: Challenges, Strategies, and Perspectives, *Adv. Funct. Mater.*, 2024, **34**, 2406639, DOI: 10.1002/adfm.202406639.

3 L. Ren, Y. Li, N. Zhang, Z. Li, X. Lin, W. Zhu, C. Lu, W. Ding and J. Zou, Nanostructuring of Mg-Based Hydrogen Storage Materials: Recent Advances for Promoting Key Applications, *Nano-Micro Lett.*, 2023, **15**, 93, DOI: 10.1007/s40820-023-01041-5.

4 X. Zhang, Y. Liu, Z. Ren, X. Zhang, J. Hu, Z. Huang, Y. Lu, M. Gao and H. Pan, Realizing 6.7 wt% reversible storage of hydrogen at ambient temperature with non-confined ultrafine magnesium hydrides, *Energy Environ. Sci.*, 2021, **14**, 2302–2313, DOI: 10.1039/D0EE03160G.

5 J. Cai, Y. Jiang, T. Yao, X. Tang, Y. Liu, Z. Xu, X. Zhao, B. Zhang, Z. Gao and W. Yang, A demand-driven dynamic heating strategy for ultrafast and energy-efficient MgH2 dehydrogenation utilizing the "burst effect,", *J. Energy Storage*, 2025, **130**, 117495, DOI: 10.1016/j.est.2025.117495.

6 J. Zhang, Z. Li, Y. Wu, X. Guo, J. Ye, B. Yuan, S. Wang and L. Jiang, Recent advances on the thermal destabilization of Mg-based hydrogen storage materials, *RSC Adv.*, 2019, **9**, 408–428, DOI: 10.1039/C8RA05596C.

7 H. Liang, W. Li and J. Zheng, *In situ* construction of dual-functional Ni/Ni x B catalysts for the hydrogenation and dehydrogenation of magnesium hydride, *J. Mater. Chem. A*, 2024, **12**, 25829–25836, DOI: 10.1039/D4TA05395H.

8 Y. Li, L. Ren, Y. Yao, Y. Zhao, H. Xu, Z. Li, Z. Li, X. Dai, Y. Tian, S. Cao, X. Lin, C. Ye, A. Züttel and J. Zou, A Single-Atom Interface Engineering Strategy to Promote Hydrogen Sorption Performances of Magnesium Hydride, *Adv. Funct. Mater.*, 2025, **35**, 2417915, DOI: 10.1002/adfm.202417915.

9 L. Ren, W. Zhu, Q. Zhang, C. Lu, F. Sun, X. Lin and J. Zou, MgH2 confinement in MOF-derived N-doped porous carbon nanofibers for enhanced hydrogen storage, *Chem. Eng. J.*, 2022, **434**, 134701, DOI: 10.1016/j.cej.2022.134701.

10 N. A. Ali, N. A. Sazelee, M. F. Md Din, M. M. Nasef, A. A. Jalil, H. Liu and M. Ismail, Ameliorating the re/dehydrogenation behaviour of MgH2 by zinc titanate addition, *J. Magnesium Alloys*, 2023, **11**, 2205–2215, DOI: 10.1016/j.jma.2023.05.005.

11 P. Meena, R. Singh, V. K. Sharma and I. P. Jain, Role of NiMn9.3Al4.0Co14.1Fe3.6 alloy on dehydrogenation kinetics of MgH2, *J. Magnesium Alloys*, 2018, **6**, 318–325, DOI: 10.1016/j.jma.2018.05.007.

12 J. A. Esterhuizen, B. R. Goldsmith and S. Linic, Interpretable machine learning for knowledge generation in heterogeneous catalysis, *Nat. Catal.*, 2022, **5**, 175–184, DOI: 10.1038/s41929-022-00744-z.

13 Z. Ma, J. Zou, D. Khan, W. Zhu, C. Hu, X. Zeng and W. Ding, Preparation and hydrogen storage properties of MgH2-trimesic acid-TM MOF (TM=Co, Fe) composites, *J. Mater. Sci. Technol.*, 2019, **35**, 2132–2143, DOI: 10.1016/j.jmst.2019.05.049.

14 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, Machine learning for molecular and materials science, *Nature*, 2018, **559**, 547–555, DOI: 10.1038/s41586-018-0337-2.

15 Y. Li, N. A. Kabanova, V. A. Blatov and J. Wang, High-throughput screening of phosphide compounds for potassium-ion conductive cathode application, *J. Mater. Inf.*, 2025, **5**, 21, DOI: 10.20517/jmi.2024.87.

16 J. Chen, Y. Zhang, J. Luan, Y. Fan, Z. Yu, B. Liu and K. Chou, Prediction of thermal conductivity in multi-component magnesium alloys based on machine learning and multiscale computation, *J. Mater. Inf.*, 2025, **5**, 22, DOI: 10.20517/jmi.2024.89.

17 S. Mooraj and W. Chen, A review on high-throughput development of high-entropy alloys by combinatorial methods, *J. Mater. Inf.*, 2023, **3**, 4, DOI: 10.20517/jmi.2022.41.

18 X. Wu, H. Wang, Y. Gong, D. Fan, P. Ding, Q. Li and Q. Qian, Graph neural networks for molecular and materials representation, *J. Mater. Inf.*, 2023, **3**, 12, DOI: 10.20517/jmi.2023.10.

19 S. Hu, X. Ma, Y. Zhang and W. Xu, Application of self-supervised learning in steel surface defect detection, *J. Mater. Inf.*, 2025, **5**, 44, DOI: 10.20517/jmi.2025.21.

20 N. S. Bobbitt and R. Q. Snurr, Molecular modelling and machine learning for high-throughput screening of metal-organic frameworks for hydrogen storage, *Mol. Simul.*, 2019, **45**, 1069–1081, DOI: 10.1080/08927022.2019.1597271.

21 J. Schmidt, M. R. G. Marques, S. Botti and M. A. L. Marques, Recent advances and applications of machine learning in solid-state materials science, *npj Comput. Mater.*, 2019, **5**, 83, DOI: 10.1038/s41524-019-0221-0.

22 T. Yao, Y. Yang, J. Cai, R. Liu, Z. Dong, X. Tang, X. Shao, Z. Gao, G. An and W. Yang, From LLM to Agent: A large-language-model-driven machine learning framework for catalyst design of MgH2 dehydrogenation, *J. Magnesium Alloys*, 2025, S2213956725002853, DOI: 10.1016/j.jma.2025.08.021.

23 X. Lu, S. Luo, J. Li, M. Chen, T. Yao, Z. Xu, Y. Yan, J. Li, X. Shao, Z. Gao and W. Yang, FIND: a forward–inverse navigation and discovery platform for hydrogen storage

alloys powered by data-driven machine learning, *J. Mater. Inf.*, 2025, **5**, 48, DOI: 10.20517/jmi.2025.56.
24 C. Chen, W. Ye, Y. Zuo, C. Zheng and S. P. Ong, Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals, *Chem. Mater.*, 2019, **31**, 3564–3572, DOI: 10.1021/acs.chemmater.9b01294.
25 M. Witman, S. Ling, D. M. Grant, G. S. Walker, S. Agarwal, V. Stavila and M. D. Allendorf, Extracting an Empirical Intermetallic Hydride Design Principle from Limited Data via Interpretable Machine Learning, *J. Phys. Chem. Lett.*, 2020, **11**, 40–47, DOI: 10.1021/acs.jpclett.9b02971.
26 M. Wang and H. Zhu, Machine Learning for Transition-Metal-Based Hydrogen Generation Electrocatalysts, *ACS Catal.*, 2021, **11**, 3930–3937, DOI: 10.1021/acscatal.1c00178.
27 M. I. Maulana Kusdhany and S. M. Lyth, New insights into hydrogen uptake on porous carbon materials via explainable machine learning, *Carbon*, 2021, **179**, 190–201, DOI: 10.1016/j.carbon.2021.04.036.
28 A. Rahnama, G. Zepon and S. Sridhar, Machine learning based prediction of metal hydrides for hydrogen storage, part II: Prediction of material class, *Int. J. Hydrogen Energy*, 2019, **44**, 7345–7353, DOI: 10.1016/j.ijhydene.2019.01.264.
29 A. Rahnama, G. Zepon and S. Sridhar, Machine learning based prediction of metal hydrides for hydrogen storage, part I: Prediction of hydrogen weight percent, *Int. J. Hydrogen Energy*, 2019, **44**, 7337–7344, DOI: 10.1016/j.ijhydene.2019.01.261.
30 J. R. Hattrick-Simpers, J. M. Gregoire and A. G. Kusne, Perspective: Composition–structure–property mapping in high-throughput experiments: Turning data into knowledge, *APL Mater.*, 2016, **4**, 053211, DOI: 10.1063/1.4950995.
31 O. Morrison, E. Uteva, G. S. Walker, D. M. Grant and S. Ling, Long Time Scale Molecular Dynamics Simulation of Magnesium Hydride Dehydrogenation Enabled by Machine Learning Interatomic Potentials, *ACS Appl. Energy Mater.*, 2025, **8**, 492–502, DOI: 10.1021/acsaem.4c02627.
32 P. Zhou, X. Xiao, X. Zhu, Y. Chen, W. Lu, M. Piao, Z. Cao, M. Lu, F. Fang, Z. Li, L. Jiang and L. Chen, Machine learning enabled customization of performance-oriented hydrogen storage materials for fuel cell systems, *Energy Storage Mater.*, 2023, **63**, 102964, DOI: 10.1016/j.ensm.2023.102964.
33 Z. Ding, Z. Chen, T. Ma, C.-T. Lu, W. Ma and L. Shaw, Predicting the hydrogen release ability of LiBH 4 -based mixtures by ensemble machine learning, *Energy Storage Mater.*, 2020, **27**, 466–477, DOI: 10.1016/j.ensm.2019.12.010.
34 Y. Sohail, C. Zhang, D. Xue, J. Zhang, D. Zhang, S. Gao, Y. Yang, X. Fan, H. Zhang, G. Liu, J. Sun and E. Ma, Machine-learning design of ductile FeNiCoAlTa alloys with high strength, *Nature*, 2025, **643**, 119–124, DOI: 10.1038/s41586-025-09160-2.
35 J.-W. Lee, W. B. Park, M. Kim, S. Pal Singh, M. Pyo and K.-S. Sohn, A data-driven XRD analysis protocol for phase identification and phase-fraction prediction of multiphase inorganic compounds, *Inorg. Chem. Front.*, 2021, **8**, 2492–2504, DOI: 10.1039/D0QI01513J.
36 F. Oviedo, Z. Ren, S. Sun, C. Settens, Z. Liu, N. T. P. Hartono, S. Ramasamy, B. L. DeCost, S. I. P. Tian, G. Romano, A. Gilad Kusne and T. Buonassisi, Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks, *npj Comput. Mater.*, 2019, **5**, 60, DOI: 10.1038/s41524-019-0196-x.
37 B. D. Lee, J.-W. Lee, J. Ahn, S. Kim, W. B. Park and K.-S. Sohn, A Deep Learning Approach to Powder X-Ray Diffraction Pattern Analysis: Addressing Generalizability and Perturbation Issues Simultaneously, *Adv. Intell. Syst.*, 2023, **5**, 2300140, DOI: 10.1002/aisy.202300140.
38 F. Oviedo, Z. Ren, S. Sun, C. Settens, Z. Liu, N. T. P. Hartono, S. Ramasamy, B. L. DeCost, S. I. P. Tian, G. Romano, A. Gilad Kusne and T. Buonassisi, Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks, *npj Comput. Mater.*, 2019, **5**, 60, DOI: 10.1038/s41524-019-0196-x.
39 B. D. Lee, J.-W. Lee, J. Ahn, S. Kim, W. B. Park and K.-S. Sohn, A Deep Learning Approach to Powder X-Ray Diffraction Pattern Analysis: Addressing Generalizability and Perturbation Issues Simultaneously, *Adv. Intell. Syst.*, 2023, **5**, 2300140, DOI: 10.1002/aisy.202300140.
40 N. J. Szymanski, S. Fu, E. Persson and G. Ceder, Integrated analysis of X-ray diffraction patterns and pair distribution functions for machine-learned phase identification, *npj Comput. Mater.*, 2024, **10**, 45, DOI: 10.1038/s41524-024-01230-9.
41 H. Dong, K. T. Butler, D. Matras, S. W. T. Price, Y. Odarchenko, R. Khatry, A. Thompson, V. Middelkoop, S. D. M. Jacques, A. M. Beale and A. Vamvakeros, A deep convolutional neural network for real-time full profile analysis of big powder diffraction data, *npj Comput. Mater.*, 2021, **7**, 74, DOI: 10.1038/s41524-021-00542-4.
42 X. Zhao, Y. Luo, J. Liu, W. Liu, K. M. Rosso, X. Guo, T. Geng, A. Li and X. Zhang, Machine Learning Automated Analysis of Enormous Synchrotron X-ray Diffraction Datasets, *J. Phys. Chem. C*, 2023, **127**, 14830–14838, DOI: 10.1021/acs.jpcc.3c03572.
43 R. Dong, Y. Zhao, Y. Song, N. Fu, S. S. Omee, S. Dey, Q. Li, L. Wei and J. Hu, DeepXRD, a Deep Learning Model for Predicting XRD spectrum from Material Composition, *ACS Appl. Mater. Interfaces*, 2022, **14**, 40102–40115, DOI: 10.1021/acsami.2c05812.
44 J. Epp, X-ray diffraction (XRD) techniques for materials characterization, in: *Materials Characterization Using Nondestructive Evaluation (NDE) Methods*, Elsevier, 2016, pp. 81–124, DOI: 10.1016/B978-0-08-100040-3.00004-3.
45 B. L. DeCost and E. A. Holm, Characterizing powder materials using keypoint-based computer vision methods, *Comput. Mater. Sci.*, 2017, **126**, 438–445, DOI: 10.1016/j.commatsci.2016.08.038.
46 D. Chen, Y. Bai, S. Ament, W. Zhao, D. Guevarra, L. Zhou, B. Selman, R. B. Van Dover, J. M. Gregoire and C. P. Gomes, Automating crystal-structure phase mapping by combining deep learning with constraint reasoning, *Nat. Mach. Intell.*, 2021, **3**, 812–822, DOI: 10.1038/s42256-021-00384-1.

47 P. Xu, X. Ji, M. Li and W. Lu, Small data machine learning in materials science, *npj Comput. Mater.*, 2023, **9**, 42, DOI: 10.1038/s41524-023-01000-z.

48 J. A. Esterhuizen, B. R. Goldsmith and S. Linic, Interpretable machine learning for knowledge generation in heterogeneous catalysis, *Nat. Catal.*, 2022, **5**, 175–184, DOI: 10.1038/s41929-022-00744-z.

49 S. Touati, A. Benghia, Z. Hebboul, I. K. Lefkaier, M. B. Kanoun and S. Goumri-Said, Machine Learning Models for Efficient Property Prediction of ABX3 Materials: A High-Throughput Approach, *ACS Omega*, 2024, **9**, 47519–47531, DOI: 10.1021/acsomega.4c06139.

50 X. Liu, T. Liu and P. Feng, Long-term performance prediction framework based on XGBoost decision tree for pultruded FRP composites exposed to water, humidity and alkaline solution, *Compos. Struct.*, 2022, **284**, 115184, DOI: 10.1016/j.compstruct.2022.115184.

51 Z. Ding, H. Li and L. Shaw, New insights into the solid-state hydrogen storage of nanostructured LiBH4-MgH2 system, *Chem. Eng. J.*, 2020, **385**, 123856, DOI: 10.1016/j.cej.2019.123856.

52 X. Zhang, X. Wang and S. Hu, Structural Fingerprinting of Crystalline Materials from XRD Patterns Using Atomic Cluster Expansion Neural Network and Atomic Cluster Expansion, *Appl. Sci.*, 2025, **15**, 5851, DOI: 10.3390/app15115851.

53 V.-A. Surdu and R. Győrgy, X-ray Diffraction Data Analysis by Machine Learning Methods—A Review, *Appl. Sci.*, 2023, **13**, 9992, DOI: 10.3390/app13179992.

54 K. Ma, Integrated hybrid modeling and SHAP (SHapley Additive exPlanations) to predict and explain the adsorption properties of thermoplastic polyurethane (TPU) porous materials, *RSC Adv.*, 2024, **14**, 10348–10357, DOI: 10.1039/D4RA00010B.

55 F. Oviedo, J. L. Ferres, T. Buonassisi and K. T. Butler, Interpretable and Explainable Machine Learning for Materials Science and Chemistry, *Acc. Mater. Res.*, 2022, **3**, 597–607, DOI: 10.1021/accountsmr.1c00244.

56 Y. Liu, H. Du, X. Zhang, Y. Yang, M. Gao and H. Pan, Superior catalytic activity derived from a two-dimensional Ti3 C2 precursor towards the hydrogen storage reaction of magnesium hydride, *Chem. Commun.*, 2016, **52**, 705–708, DOI: 10.1039/C5CC08801A.

57 S. Wang, M. Gao, Z. Yao, K. Xian, M. Wu, Y. Liu, W. Sun and H. Pan, High-loading, ultrafine Ni nanoparticles dispersed on porous hollow carbon nanospheres for fast (de) hydrogenation kinetics of MgH2, *J. Magnesium Alloys*, 2022, **10**, 3354–3366, DOI: 10.1016/j.jma.2021.05.004.

58 K. Xian, M. Wu, M. Gao, S. Wang, Z. Li, P. Gao, Z. Yao, Y. Liu, W. Sun and H. Pan, A Unique Nanoflake-Shape Bimetallic Ti–Nb Oxide of Superior Catalytic Effect for Hydrogen Storage of MgH2, *Small*, 2022, **18**, 2107013, DOI: 10.1002/smll.202107013.

59 X. Wang, M. Liu, T. Tian, F. Liu, J. Wang, J. Li and G. Liu, Fibrous V/Nb bimetallic oxides with remarkable catalytic effect on hydrogen storage properties of MgH2, *Int. J. Hydrogen Energy*, 2025, **97**, 1168–1176, DOI: 10.1016/j.ijhydene.2024.12.021.

60 Z. Wang, X. Zhang, Z. Ren, Y. Liu, J. Hu, H. Li, M. Gao, H. Pan and Y. Liu, *In situ* formed ultrafine NbTi nanocrystals from a NbTiC solid-solution MXene for hydrogen storage in MgH2, *J. Mater. Chem. A*, 2019, **7**, 14244–14252, DOI: 10.1039/C9TA03665B.

61 R. R. Shahi, A. P. Tiwari, M. A. Shaz and O. N. Srivastava, Studies on de/rehydrogenation characteristics of nanocrystalline MgH2 co-catalyzed with Ti, Fe and Ni, *Int. J. Hydrogen Energy*, 2013, **38**, 2778–2784, DOI: 10.1016/j.ijhydene.2012.11.073.

62 Y. Chen, B. Sun, G. Zhang, Y. Wang and X. Li, Catalytic effect of double transition metal sulfide NiCo2S4 on hydrogen storage properties of MgH2, *Appl. Surf. Sci.*, 2024, **645**, 158801, DOI: 10.1016/j.apsusc.2023.158801.

63 Y. Jiang, N. Si, Z. Wang, H. Zhang and W. Jiang, Improved Hydrogen Storage Kinetic Properties of MgH2 with NiO/NiCo(Fe)2 O4/(Ni), *Energy Fuels*, 2024, **38**, 23804–23814, DOI: 10.1021/acs.energyfuels.4c04640.

64 W. Li, X. Yang, Q. Hou, X. Lu, J. Kong and J. Su, Modification of MgH2 hydrogen storage performance by nickel-based composite catalyst Ni/NiO, *Heliyon*, 2024, **10**, e30688, DOI: 10.1016/j.heliyon.2024.e30688.

65 L. Ji, L. Zhang, X. Yang, X. Zhu and L. Chen, The remarkably improved hydrogen storage performance of MgH2 by the synergetic effect of an FeNi/rGO nanocomposite, *Dalton Trans.*, 2020, **49**, 4146–4154, DOI: 10.1039/D0DT00230E.

66 F. U. Kosasih, F. Su, T. Du, S. R. Ratnasingham, J. Briscoe and C. Ducati, Deep Learning-Assisted Multivariate Analysis for Nanoscale Characterization of Heterogeneous Beam-Sensitive Materials, *Microsc. Microanal.*, 2023, **29**, 1047–1061, DOI: 10.1093/micmic/ozad033.

67 V. Gupta, S. K. Gupta and A. Shetty, Fractal-based supervised approach for dimensionality reduction of hyperspectral images, *Comput. Geosci.*, 2024, **193**, 105733, DOI: 10.1016/j.cageo.2024.105733.

68 Z. Ding, Z. Chen, T. Ma, C.-T. Lu, W. Ma and L. Shaw, Predicting the hydrogen release ability of LiBH 4 -based mixtures by ensemble machine learning, *Energy Storage Mater.*, 2020, **27**, 466–477, DOI: 10.1016/j.ensm.2019.12.010.

69 O. A. Bulavchenko and Z. S. Vinokurov, In Situ X-ray Diffraction as a Basic Tool to Study Oxide and Metal Oxide Catalysts, *Catalysts*, 2023, **13**, 1421, DOI: 10.3390/catal13111421.

70 C. Duan, Y. Tian, X. Wang, M. Wu, D. Fu, Y. Zhang, W. Lv, Z. Su, Z. Xue and Y. Wu, Ni-CNTs as an efficient confining framework and catalyst for improving dehydriding/rehydriding properties of MgH2, *Renewable Energy*, 2022, **187**, 417–427, DOI: 10.1016/j.renene.2022.01.048.

71 H. Zhang, Y. Bu, W. Xiong, K. He, T. Yu, Z. Liu, J. Wu, B. Li, L. Sun, Y. Zou, W. Sun, R. Guo, F. Xu, K. Zhang and H. Zhang, Effect of bimetallic nitride NiCoN on the hydrogen absorption and desorption properties of MgH2 and the catalytic effect of in situ formed Mg2Ni and Mg2Co phases, *J. Alloys Compd.*, 2023, **965**, 171431, DOI: 10.1016/j.jallcom.2023.171431.

72 M. Chen, X. Xiao, M. Zhang, M. Liu, X. Huang, J. Zheng, Y. Zhang, L. Jiang and L. Chen, Excellent synergistic

catalytic mechanism of in-situ formed nanosized Mg2Ni and multiple valence titanium for improved hydrogen desorption properties of magnesium hydride, *Int. J. Hydrogen Energy*, 2019, **44**, 1750–1759, DOI: 10.1016/j.ijhydene.2018.11.118.

73 H. Liu, C. Lu, X. Wang, L. Xu, X. Huang, X. Wang, H. Ning, Z. Lan and J. Guo, Combinations of V2 C and Ti3 C2 MXenes for Boosting the Hydrogen Storage Performances of MgH2, *ACS Appl. Mater. Interfaces*, 2021, **13**, 13235–13247, DOI: 10.1021/acsami.0c23150.

74 L. Zhang, L. Chen, X. Fan, X. Xiao, J. Zheng and X. Huang, Enhanced hydrogen storage properties of MgH2 with numerous hydrogen diffusion channels provided by Na2 Ti3 O7 nanotubes, *J. Mater. Chem. A*, 2017, **5**, 6178–6185, DOI: 10.1039/C7TA00566K.

75 Z. Yu, W. Zhang, Y. Zhang, Y. Fu, Y. Cheng, S. Guo, Y. Li and S. Han, Remarkable kinetics of novel Ni@CeO2–MgH2 hydrogen storage composite, *Int. J. Hydrogen Energy*, 2022, **47**, 35352–35364, DOI: 10.1016/j.ijhydene.2022.08.121.

76 P. Xiao, J. Liu, D. Guo, L. Yang, L. Sun, S. Li, L. Xu and H. Liu, High catalytic activity derived from TiNbAlC MAX towards improving the hydrogen storage properties of MgH2, *J. Alloys Compd.*, 2023, **955**, 170297, DOI: 10.1016/j.jallcom.2023.170297.

77 Z. Ma, J. Zhang, Y. Zhu, H. Lin, Y. Liu, Y. Zhang, D. Zhu and L. Li, Facile Synthesis of Carbon Supported Nano-Ni Particles with Superior Catalytic Effect on Hydrogen Storage Kinetics of MgH2, *ACS Appl. Energy Mater.*, 2018, **1**, 1158–1165, DOI: 10.1021/acsaem.7b00266.

78 Q. Hou, X. Yang, J. Zhang, W. Yang and E. Lv, Catalytic effect of NiO/C derived from Ni-UMOFNs on the hydrogen storage performance of magnesium hydride, *J. Alloys Compd.*, 2022, **899**, 163314, DOI: 10.1016/j.jallcom.2021.163314.

79 H. Leng, N. Miao and Q. Li, Improved hydrogen storage properties of MgH2 by the addition of KOH and graphene, *Int. J. Hydrogen Energy*, 2020, **45**, 28183–28189, DOI: 10.1016/j.ijhydene.2020.03.070.

80 T. Taufik, Z. A. Jalil and M. Mursal, Use of SiO2 Rice Husk Ash and Ni in Materials Solid Hydrogen Storage Based on MgH2, *J. Aceh Phys. Soc.*, 2020, **9**, 9–12, DOI: 10.24815/jacps.v9i1.15047.

81 M. Maulinda, Z. Jalil and A. Rahwanto, The Effect Of Magnetite (Fe3o4)Catalyst From Iron Sands On Desorption Temperature Of Mgh2 Hydrogen Storage Material, *J. Nat.*, 2016, **16**, 18–20, DOI: 10.24815/jn.v16i1.4362.

82 M. Zhang, X. Xiao, X. Wang, M. Chen, Y. Lu, M. Liu and L. Chen, Excellent catalysis of TiO2 nanosheets with high-surface-energy {001} facets on the hydrogen storage properties of MgH2, *Nanoscale*, 2019, **11**, 7465–7473, DOI: 10.1039/C8NR10275A.

83 C. An and Q. Deng, Improvement of Hydrogen Desorption Characteristics of MgH2 With Core-shell Ni@C Composites, *Molecules*, 2018, **23**, 3113, DOI: 10.3390/molecules23123113.

84 R. Floriano, D. R. Leiva, J. G. Dessi, A. A. C. Asselli, A. M. Jorge Junior and W. J. Botta, Mg-based Nanocomposites for Hydrogen Storage Containing Ti-Cr-V Alloys as Additives, *Mat. Res.*, 2016, **19**, 80–85, DOI: 10.1590/1980-5373-mr-2016-0179.

85 J. Li, S. Wang, Y. Du and W. Liao, Catalytic effect of Ti2C MXene on the dehydrogenation of MgH2, *Int. J. Hydrogen Energy*, 2019, **44**, 6787–6794, DOI: 10.1016/j.ijhydene.2019.01.189.

86 L. Zhang, Z. Cai, X. Zhu, Z. Yao, Z. Sun, L. Ji, N. Yan, B. Xiao and L. Chen, Two-dimensional ZrCo nanosheets as highly effective catalyst for hydrogen storage in MgH2, *J. Alloys Compd.*, 2019, **805**, 295–302, DOI: 10.1016/j.jallcom.2019.07.085.

87 M. Wu, M. Gao, S. Qu, Y. Liu, W. Sun, X. Zhang, C. Liang, X. Zhang, Y. Yang and H. Pan, LaVO4 prepared by a high–yield method for superior catalysis to the hydrogen storage of MgH2, *J. Magnesium Alloys*, 2025, **13**, 613–625, DOI: 10.1016/j.jma.2024.03.025.

88 M. Chen, X. Xiao, X. Wang, Y. Lu, M. Zhang, J. Zheng and L. Chen, Self-templated carbon enhancing catalytic effect of ZrO2 nanoparticles on the excellent dehydrogenation kinetics of MgH2, *Carbon*, 2020, **166**, 46–55, DOI: 10.1016/j.carbon.2020.05.025.

89 M. Zhang, X. Xiao, Z. Hang, M. Chen, X. Wang, N. Zhang and L. Chen, Superior catalysis of NbN nanoparticles with intrinsic multiple valence on reversible hydrogen storage properties of magnesium hydride, *Int. J. Hydrogen Energy*, 2021, **46**, 814–822, DOI: 10.1016/j.ijhydene.2020.09.173.

90 Z. Wang, Z. Ren, N. Jian, M. Gao, J. Hu, F. Du, H. Pan and Y. Liu, Vanadium oxide nanoparticles supported on cubic carbon nanoboxes as highly active catalyst precursors for hydrogen storage in MgH2, *J. Mater. Chem. A*, 2018, **6**, 16177–16185, DOI: 10.1039/C8TA05437A.

91 B. Rathi, S. Agarwal, K. Shrivastava, H. Miyaoka, T. Ichikawa, M. Kumar and A. Jain, Understanding the kinetic enhancement of hydrogen sorption in MgH2 via KNbO3 through the perspective of the catalytic mechanism, *Int. J. Hydrogen Energy*, 2025, **120**, 213–224, DOI: 10.1016/j.ijhydene.2025.03.273.

92 Y. Pan, X. Yang, X. Lu, J. Su, J. Kong, D. Huo, W. Zhang and Q. Hou, Enhancement catalysis of layered CuMoO4 on hydrogen storage performance of MgH2, *J. Alloys Compd.*, 2025, **1017**, 179129, DOI: 10.1016/j.jallcom.2025.179129.

93 Y. Liu, Y. Zhu, Q. Tang, X. Shen, Y. Zhu, J. Zhang, J. Wang, Z. Ba, W. Cai, Y. Huang and X. Hu, In situ formation of highly dispersed LaH3 and Mg2Ni derived from 3DOM LaNiO3 for the hydrogen storage kinetics enhancement of MgH2, *J. Alloys Compd.*, 2025, **1032**, 181069, DOI: 10.1016/j.jallcom.2025.181069.

94 M. Wu, M. Gao, S. Qu, Y. Liu, W. Sun, X. Zhang, C. Liang, X. Zhang, Y. Yang and H. Pan, LaVO4 prepared by a high–yield method for superior catalysis to the hydrogen storage of MgH2, *J. Magnesium Alloys*, 2025, **13**, 613–625, DOI: 10.1016/j.jma.2024.03.025.

95 S. Kumar, Y. Kojima and G. K. Dey, Morphological effects of Nb2O5 on Mg–MgH2 system for thermal energy storage application, *Int. J. Hydrogen Energy*, 2018, **43**, 809–816, DOI: 10.1016/j.ijhydene.2017.11.039.