

# Machine Learning in Solid-State Hydrogen Storage Materials: Challenges and Perspectives

Panpan Zhou, Qianwen Zhou, Xuezhang Xiao,\* Xiulin Fan, Yongjin Zou, Lixian Sun, Jinghua Jiang, Dan Song, and Lixin Chen\*

Machine learning (ML) has emerged as a pioneering tool in advancing the research application of high-performance solid-state hydrogen storage materials (HSMs). This review summarizes the state-of-the-art research of ML in resolving crucial issues such as low hydrogen storage capacity and unfavorable de-/hydrogenation cycling conditions. First, the datasets, feature descriptors, and prevalent ML models tailored for HSMs are described. Specific examples include the successful application of ML in titanium-based, rare-earth-based, solid solution, magnesium-based, and complex HSMs, showcasing its role in exploiting composition–structure–property relationships and designing novel HSMs for specific applications. One of the representative ML works is the single-phase Ti-based HSM with superior cost-effective and comprehensive properties, tailored to fuel cell hydrogen feeding system at ambient temperature and pressure through high-throughput composition–performance scanning. More importantly, this review also identifies and critically analyzes the key challenges faced by ML in this domain, including poor data quality and availability, and the balance between model interpretability and accuracy, together with feasible countermeasures suggested to ameliorate these problems. In summary, this work outlines a roadmap for enhancing ML’s utilization in solid-state hydrogen storage research, promoting more efficient and sustainable energy storage solutions.

Among various clean energy sources, hydrogen energy has been hailed as one of the ideal energy carriers due to its high energy density, abundant resources, clean, and pollution-free attributes.<sup>[1]</sup> Yet, some technical barriers severely limit the roll-out of a hydrogen-powered economy, including the safe and efficient storage and transport of hydrogen energy. Solid-state hydrogen storage technologies are favored over the traditional high-pressure gaseous hydrogen storage and low-temperature liquid hydrogen storage technologies with advantages of high volumetric hydrogen storage capacity, satisfactory reversible de-/hydrogenation performance, excellent safety feature, and practicality.<sup>[2]</sup>

To satisfy the growing demand for solid-state hydrogen storage and supply, the urgent need arises for the efficient generation of high-performance solid-state HSMs with enhanced saturated hydrogen storage capacity, higher reversible hydrogen storage efficiency, milder de-/hydrogenation conditions, improved cycling durability, and safer operation. Traditional trial-and-error experimental methods have proven to be inefficient in keeping pace with the rigorous

experimental conditions (de-/hydrogenation pressure and temperature) and the accelerated advancements in high-performance materials. Since the 1980s, the convergence of materials science, physics, and computational science has significantly accelerated the evolution of diverse computational simulation techniques

## 1. Introduction

Extreme weather caused by the greenhouse effect has led to the recognition of the urgency of advancing the energy revolution and vigorously developing green and renewable energy sources.

P. Zhou, J. Jiang, D. Song  
College of Materials Science and Engineering

Hohai University  
Changzhou, Jiangsu 213200, China

P. Zhou, Q. Zhou, X. Xiao, X. Fan, L. Chen  
State Key Laboratory of Silicon and Advanced Semiconductor Materials  
School of Materials Science and Engineering  
Zhejiang University  
Hangzhou, Zhejiang 310058, China  
E-mail: [zxiao@zju.edu.cn](mailto:zxiao@zju.edu.cn); [lxchen@zju.edu.cn](mailto:lxchen@zju.edu.cn)

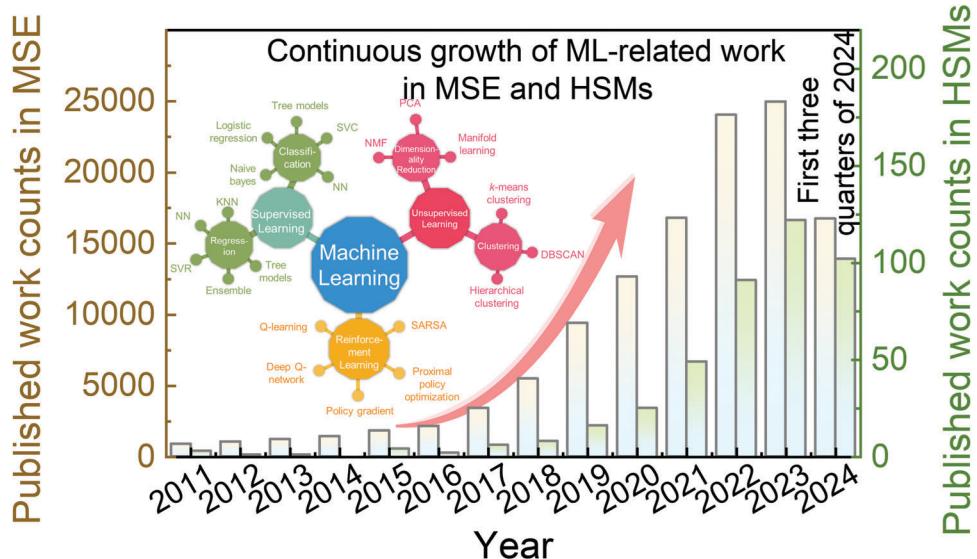
X. Xiao  
School of Advanced Energy  
Sun Yat-Sen University  
Shenzhen 518107, China

Y. Zou, L. Sun  
Guangxi Key Laboratory of Information Materials  
Guilin University of Electronic Technology  
Guilin 541004, China

L. Chen  
Key Laboratory of Hydrogen Storage and Transportation Technology of  
Zhejiang Province  
Hangzhou, Zhejiang 310027, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adma.202413430>

DOI: 10.1002/adma.202413430



**Figure 1.** Published work counts from 2011 to 2024 found by searching on the database “Web of Science” with “Machine Learning” + “Materials” as well as “Machine Learning” + “Hydrogen Storage Materials” as keywords.

and the exploration of solid-state HSMs, specifically including density functional theory (DFT), ab initio molecular dynamics simulation (AIMD), finite element simulation (FE), etc. For example, based on DFT calculations, Hu et al.<sup>[3]</sup> discovered that the transition of TiZrHfMoNb high-entropy HSM from BCC to FCC phases stems from the disordering of the H-H bonds and the negative effect of structural stabilization resulting from the increase of the hydrogen content, and that the hydrogen atoms basically occupy the tetrahedral interstitial sites at H/M = 2. Dong et al.<sup>[4]</sup> revealed that the heterojunction system formed by single-atom catalysts containing different transition metals and MgH<sub>2</sub> can effectively reduce the dehydrogenation energy barrier of MgH<sub>2</sub> by means of DFT and AIMD, and the system electronegativity of SACs can be used as an effective descriptor for predicting the dehydrogenation energy barrier. Zhan et al.<sup>[5]</sup> mixed high-density HSM powders, curable silica gel, and graphite, and pressed them into a block structure, effectively mitigating the stress concentration potential caused by the localized powder accumulation. Through FE simulation, it's concluded that incorporating high thermal conductivity graphite can reduce the total de-/hydrogenation time to about 40% of that required by the original metal hydride bed. As previously elaborated, computational simulation methods have achieved remarkable success in multi-scale research on solid-state HSMs, demonstrating their effectiveness and versatility in this field.

Given the cognitive limitations of current research on the storage processes and forms of hydrogen atoms, both experimental measurements and computational simulations face constraints in effectively exploring the intricate and vast chemical space of solid-state HSMs. This often leads to substantial resource expenditure and the occurrence of undesirable negative feedback outcomes. The primary goal of Materials Genetic Engineering (MGI),<sup>[6]</sup> proposed in 2011, is to expedite the transformation of the materials research and development (*R & D*) model through experimental, computational, and theoretical high-throughput integrated innovation, thereby significantly reducing the *R&D* cy-

cle time, lowering the *R&D* costs, and vigorously promoting the development and engineering application of new materials.<sup>[7]</sup> As exhibited in **Figure 1**, based on literature search results from the Web of Science database, the digitally-driven materials research paradigm has entered an exceptionally rapid development stage and has been successfully implemented in structural materials,<sup>[8]</sup> catalyst materials,<sup>[9]</sup> battery materials,<sup>[10]</sup> and hydrogen storage materials,<sup>[11]</sup> among others.

Several excellent review articles on materials informatics have been published,<sup>[12]</sup> however, the summary of the latest research progresses and perspectives of ML in the field of solid-state hydrogen storage have not been reported. In this review, not only do we analyze ML framework applicable to the development of solid-state HSMs, but also systematically summarize the current situations, primary limitations, and potential perspectives of ML in the field of solid-state HSMs, with a view to broadening the horizons for customization of high-performance solid-state HSMs. The remainder of this review is structured as follows: Section 2 briefly introduces the fundamental concepts and workflows of ML. Then, Sections 3 and 4 present the state-of-the-art of ML research and challenges in the field of solid-state hydrogen storage, respectively. Finally, the summary and prospects are given in Section 5.

## 2. Machine Learning Processes and Related Concepts

ML is an advanced branch of artificial intelligence that grants computer systems the power to learn from data and make predictions or decisions to accomplish specific tasks.<sup>[13]</sup> Therefore, the exploration of ML represents a promising avenue for the effective development of high-performance solid-state HSMs. **Figure 2** displays basic ML processes for materials science, consisting of dataset establishment, feature engineering, model training, and model performance evaluation and application.<sup>[14]</sup> Firstly, the primary objectives to be addressed by ML should

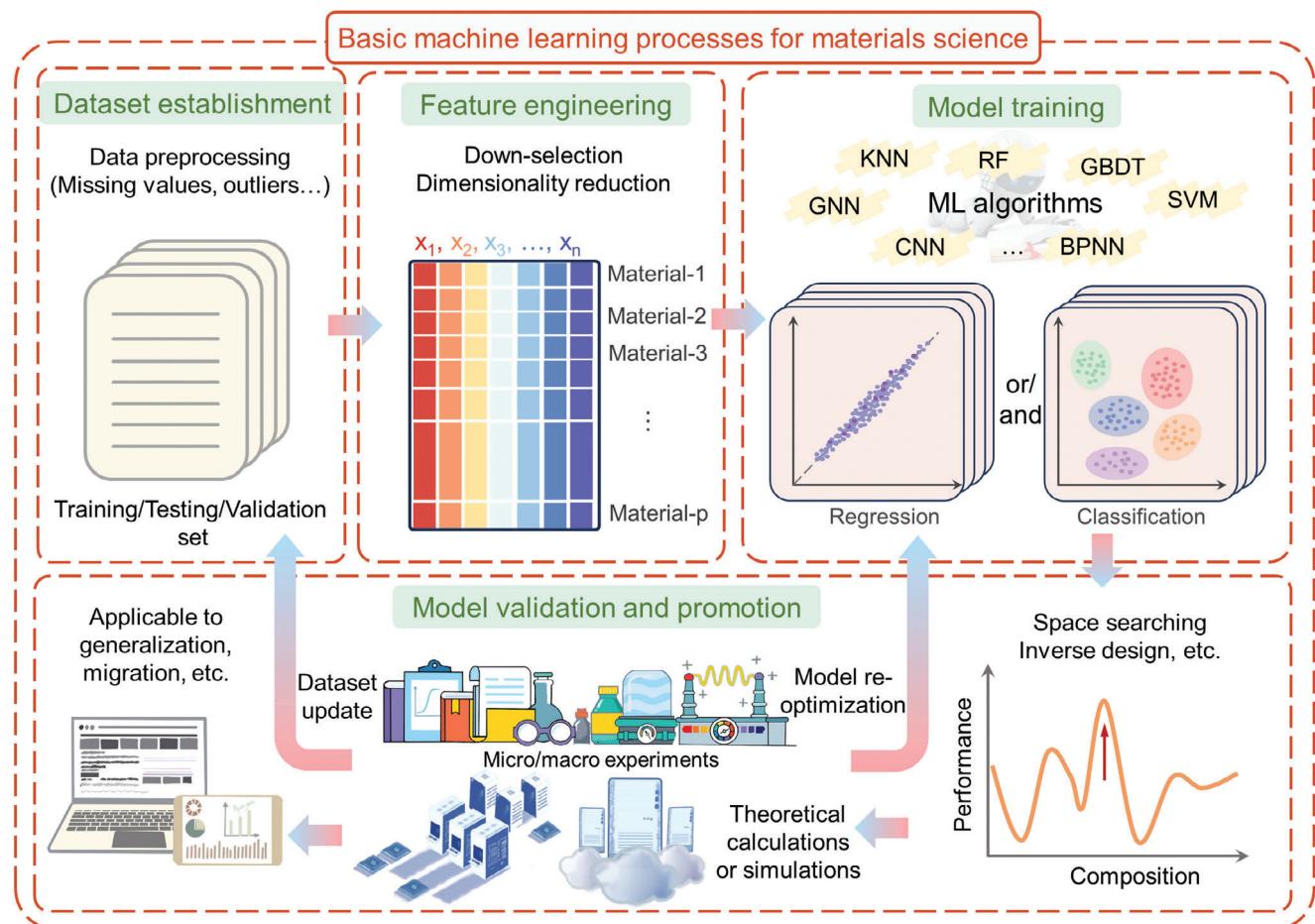


Figure 2. The basic machine learning processes for materials science.

be identified (e.g., optimization of phase structure, modulation of de-/hydrogenation performance, enhancement of hydrogen storage capacity, etc.). Following this, the establishment and preprocessing of specialized datasets, along with the design and selection of feature descriptors for solid-state HSMs, are pivotal for ensuring optimal model training and prediction performance. These steps will be elaborated in Sections 2.1 and 2.2. Subsequently, suitable ML models can be chosen based on the specific nature of the problem, with parameters fine-tuned to accurately predict or classify new data. Sections 2.3 and 2.4 will delve into the principles, strengths, and weaknesses of common ML models, along with an introduction to the main evaluation metrics. Additionally, it is crucial to guard against common pitfalls such as overfitting (characterized by high training accuracy but low prediction accuracy) and underfitting (low training and prediction accuracy). These issues can be mitigated by improving data quality and quantity, carefully selecting the model and tuning hyper-parameters, applying regularization techniques, etc. Ultimately, the trained model will be deployed to solve the specific problems through prediction or classification, with necessary adjustments and optimizations made based on practical application outcomes.

## 2.1. Datasets for Hydrogen Storage Materials

In the digital-driven paradigm for materials research, the quality of the dataset and the information in feature descriptors set the upper boundary for the accuracy of ML models, while the optimization of models and algorithms merely aims to approach this limit as closely as possible.<sup>[15]</sup> Thus, the establishment of the informative dataset and feature descriptors suitable for solid-state HSMs is the priority issue to be handled.<sup>[16]</sup>

Currently, well-known databases for inorganic materials include Materials Project (MP),<sup>[17]</sup> Crystallography Open Database (COD),<sup>[18]</sup> Inorganic Crystal Structure Database (ICSD),<sup>[19]</sup> Open Quantum Materials Database (OQMD),<sup>[20]</sup> Novel Materials Discovery (NOMAD),<sup>[21]</sup> Automatic Flow for Materials Discovery (AFLOWLIB),<sup>[22]</sup> Pauling file,<sup>[23]</sup> NIST,<sup>[24]</sup> among others. These experimentally or computationally based databases directly provide information on the crystallographic structure, physicochemical properties, and electronic structure of inorganic materials. However, the aforesaid databases do not include data on the structural and thermodynamic hydrogen storage performance of HSMs, and therefore cannot be directly applied to the research and development of solid-state HSMs.

**Table 1.** Some available and representative datasets for solid-state hydrogen storage materials.

Materials class	Data entries	Materials properties	Source of data	Refs.
Multi-category HSMs (HydPARK)	2722	1) Hydrogen storage capacity 2) Heat of formation	<a href="https://datahub.hymarc.org/dataset/hydrogen-storage-materials-db/resource/775e172f-9827-44bb-aebe-f5ae488d0419">https://datahub.hymarc.org/dataset/hydrogen-storage-materials-db/resource/775e172f-9827-44bb-aebe-f5ae488d0419</a>	[25]
Multi-category HSMs (ML-HydPARK v0.0.0)	398	1) Hydrogen storage capacity 2) Heat of formation 3) Entropy of formation 4) Equilibrium pressure	<a href="https://zenodo.org/records/7293045">https://zenodo.org/records/7293045</a>	[27c,28]
Multi-category HSMs (ML-HydPARK v0.0.1)	400	1) Hydrogen storage capacity 2) Heat of formation 3) Entropy of formation 4) Equilibrium pressure	<a href="https://zenodo.org/records/7293161">https://zenodo.org/records/7293161</a>	[27c,28]
Multi-category HSMs (ML-HydPARK v0.0.2)	428	1) Hydrogen storage capacity 2) Heat of formation 3) Entropy of formation 4) Equilibrium pressure	<a href="https://zenodo.org/records/7293164">https://zenodo.org/records/7293164</a>	[27c,28]
Multi-category HSMs (ML-HydPARK v0.0.3)	429	1) Hydrogen storage capacity 2) Heat of formation 3) Entropy of formation 4) Equilibrium pressure	<a href="https://zenodo.org/records/7324807">https://zenodo.org/records/7324807</a>	[27c,28]
Multi-category HSMs (ML-HydPARK v0.0.4)	430	1) Hydrogen storage capacity 2) Heat of formation 3) Entropy of formation 4) Equilibrium pressure	<a href="https://zenodo.org/records/7324809">https://zenodo.org/records/7324809</a>	[27c,28]
Multi-category HSMs (ML-HydPARK v0.0.5)	772	1) Hydrogen storage capacity 2) Heat of formation 3) Entropy of formation 4) Equilibrium pressure	<a href="https://zenodo.org/records/10680097">https://zenodo.org/records/10680097</a>	[27c,28]
Complex and high-density HSMs	2150	1) Heat of formation 2) Phase abundance 3) Hydrogen storage capacity	<a href="http://172.25.0.34/portal/index.html">http://172.25.0.34/portal/index.html</a>	[29]
AB <sub>2</sub> -type HSMs	314	1) Hydride formation enthalpy 2) Hydride formation entropy 3) Phase abundance 4) Hydrogen storage capacity	<a href="https://ars.els-cdn.com/content/image/1-s2.0-S0360319922004517-mmc1.xlsx">https://ars.els-cdn.com/content/image/1-s2.0-S0360319922004517-mmc1.xlsx</a>	[32]
AB <sub>2</sub> -type HSMs	420	Hydride formation enthalpy	<a href="https://www.sciencedirect.com/science/article/pii/S1385894724040932#s0065">https://www.sciencedirect.com/science/article/pii/S1385894724040932#s0065</a>	[33]
Mg-based HSMs (DFT data)	15 295	Formation energy	<a href="https://data.mendeley.com/datasets/9mn3ptmnb2/1">https://data.mendeley.com/datasets/9mn3ptmnb2/1</a>	[34]
Mg-based HSMs	636	1) Heat of hydride formation 2) Equilibrium potential	<a href="https://ars.els-cdn.com/content/image/1-s2.0-S2352152X23011179-mmc2.csv">https://ars.els-cdn.com/content/image/1-s2.0-S2352152X23011179-mmc2.csv</a>	[35]

At present, datasets targeted at solid-state HSMs fall into two main categories, the HydPARK dataset proposed by the U.S. Department of Energy (DOE),<sup>[25]</sup> and proprietary datasets constructed for specific applications, as presented in **Table 1**. HydPARK dataset covers A<sub>2</sub>B-type HSMs, AB-type HSMs, AB<sub>2</sub>-type HSMs, AB<sub>5</sub>-type HSMs, solid solution-type HSMs, Mg-based HSMs, complex hydrides and other intermetallic compounds that are reported from 1949 to 2002 with a total of 2722 data entries, where A denotes strong hydrogen-affinity elements (e.g., Rare earth elements, Li, Mg, Ca, Ti, Zr, V, Nb, Hf, etc.), and B denotes weak hydrogen-affinity elements with catalytic hydrogen dissociation ability (e.g., Al, Pd, V, Cr, Mn, Fe, Co, Ni, Cu, Mo etc.).<sup>[26]</sup> Nevertheless, this dataset suffers from serious overlapping information and missing data issues, so it always requires further data preprocessing.<sup>[25a,27]</sup> Accordingly, several high-quality versions of ML-HydPARK datasets were established and updated.<sup>[27c,28]</sup> ML research based on these datasets focuses

on an in-depth analysis of the hydrogen storage performance and common intrinsic patterns among different HSMs, which will be presented in Section 3.1. In contrast, the construction of the other category of HSM datasets generally requires data mining or literature screening for a specific HSM system, providing direct guidance for mechanism mining and high-throughput performance prediction of that specific hydrogen storage system. A typical example of the proprietary datasets is the one established by Huang et al.<sup>[29]</sup> for complex and high-density HSMs, which contain more than 2000 data entries such as structural information, de-/hydrogenation behavior pathways, hydrogen storage capacity, electronic structure state, etc. However, the existing HSM datasets based on experimental data suffer from small scale and poor data quality, and are mainly related to the experimental testing conditions. Specific descriptions and advanced algorithms/techniques for small datasets will be elaborated in Section 4.1.

Before the adoption of the selected or constructed dataset for subsequent ML, data cleaning and pre-processing processes are indispensable. A preliminary understanding of the data distribution, trends, missing values, outliers, etc. is foremost needed with the assistance of exploratory data analysis, and necessary experimental validation and supplementation or outright deletion of anomalous data will be carried out based on professional expertise.<sup>[27c,f,30]</sup> Afterwards, the dataset is reasonably partitioned into training, testing and validation sets, enabling it to offer considerable prediction ability even on unseen data.<sup>[30a,31]</sup>

## 2.2. Feature Descriptors for Hydrogen Storage Materials

### 2.2.1. Feature Design

The design and selection of simple and informative features constitute an important support for the implementation of predictable ML models. Based on a given HSM, the potential features reported in literature include Magpie framework,<sup>[36]</sup> structure-based framework, fundamental properties of elements, composition-related properties, hydrogenation properties, and de-/hydrogenation conditions, as documented in **Table 2**. In general, the establishment of these HSM-features relies on the composition, structure, underlying physicochemical properties, configuration states, elemental hydrogenation characteristics, and de-/hydrogenation conditions. It is worth noting that acquiring all these features is relatively straightforward.

When data size and the amount of feature information permitted, composition- or structure-based feature frameworks may qualify as an effective ML strategy. Magpie features proposed by Ward et al.<sup>[36]</sup> constitute a powerful feature framework for analyzing and predicting properties of inorganic materials. According to the code provided,<sup>[36]</sup> a determined material composition corresponds to 145 Magpie features, which can be divided into four sub-categories, including stoichiometric features (depending on the percentage of elements), elemental property-based features (multi-faceted elemental information statistics), electronic structure features (depending on the distribution of *s*, *p*, *d*, and *f* orbital electrons outside the nucleus), and ionization characteristic-based features (determining the ionic properties of the material). As for the structure-based framework, it can be generated by geometrical structure and graph neural network (GNN), including nodes, edges, and graphs.<sup>[37]</sup> In specific, nodes represent the fundamental units of material structure and may consist of atom-type, degree, hybridization, charge, etc. Edges, in turn, signify the connectivity between these nodes, mirroring chemical bonds or other interactions in the structure of the material, and may involve bond-type, stereo, etc. Regarding graph-level features, they can be extracted by specialized graph pooling or graph readout operations, covering aspects like the weight of the graph, the total number of bonds, etc.

However, the feature frameworks (Magpie-based and structure-based frameworks) serve as general-purpose tools in inorganic materials research, facilitating model training across various material systems with specific target values. Unfortunately, in many cases, these frameworks fail to establish a robust connection with the target values of the solid-state HSMs being investigated, thereby limiting their interpretability in

exploring the composition-structure-performance relationship. Consequently, to enhance the interpretability of ML models, researchers oftentimes conduct ML model training based on the feature descriptors that are established based on the underlying hydrogenation characteristics of the constituent elements themselves or other operation conditions. Therefore, it is necessary to consider and construct multi-category features in practice, among which the most prominent ones are the fundamental properties of elements, including elemental information, geometric/electronic structure information, and physicochemical information.<sup>[11d,e,38]</sup> Derived from Hume-Rothery rules and configuration entropy effect, features such as lattice distortion, mixing entropy/enthalpy, and average valence electron concentration are critical, especially for multi-principal element materials.<sup>[11e,39]</sup> Since de-/hydrogenation cycle of solid-state HSMs is a reversible process of hydrogen insertion and detachment in tetrahedral or octahedral interstitial sites, it's necessary to acquire the intrinsic hydrogen affinity factor, energy/volume altering with phase transformation, *d*-band center, highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), etc.<sup>[11d,e,40]</sup> Besides, de-/hydrogenation conditions can directly determine the manifestation of hydrogen storage performance, thus features such as atmosphere, temperature, pressure, time, catalyst-based factors, device shape-based descriptions, etc. necessitate their consideration.<sup>[11e,41]</sup>

### 2.2.2. Feature Selection

Common feature variables belonging to diverse categories are detailedly described above. Sometimes, the features are built in a way that considers the mean and variance of the features or the combination between different features. However, multicollinearity among the initially constructed or selected features can impede the training process and compromise the generalization ability of the model. Thus, the selection of the most appropriate features is essential and necessary. There are two main feature selection strategies, i.e., down-selection and dimensionality reduction.<sup>[43]</sup>

Down-selection refers to the process of selecting a significantly reduced yet highly representative set of features, and generally includes univariate statistics, model-based selection, and iterative selection.<sup>[30a,32,44]</sup> In univariate statistics, the features with the highest confidence level will be selected by assessing the statistical significance of the correlation between each individual feature and the target value.<sup>[45]</sup> Model-based feature selection employs supervised ML models to measure the importance of each individual feature and retains only the most critical features. Examples of such methods include the sparse coefficients obtained from LASSO model, and feature importance analysis based on tree models.<sup>[46]</sup> For example, Wexler et al.<sup>[47]</sup> employed regularized random forest to extract ten most important features from 29 initially considered candidates for activity discovery. As for iterative selection done with ML models, there are two approaches: starting from zero and gradually adding features until the termination condition is met, and starting with all features and individually removing them until the target threshold is reached.<sup>[30a,48]</sup> One representative approach is recursive feature elimination (RFE). Zhou et al.<sup>[11d]</sup> utilized RFE with random forest (RF) as the base estimator to successfully select 20 most

**Table 2.** Feature descriptors suitable for solid-state hydrogen storage materials.

Feature category	Feature description	Refs.
Magpie framework (145)	1) Stoichiometric-based features 2) Elemental property-based features 3) Electronic structure features 4) Ionization characteristic-based features	[30a]
Structure-based framework	1) Nodes: atom-type, degree, hybridization, charge, etc. 2) Edges: bond-type, stereo, etc. 3) Graph: weight, bonds, etc.	[37a–c,42]
Fundamental properties of elements	1) Elemental composition 2) Atomic number 3) Atomic radius 4) Atomic mass 5) Atomic density 6) Lattice constant 7) Electron affinity energy 8) Ionization potential 9) Bulk modulus 10) Shear modulus 11) Poisson's ratio 12) Melting point 13) Boiling point ...	[11d,e,38]
Composition-related properties	1) Lattice distortion 2) Mixing enthalpy 3) Mixing entropy 4) Average valence electron concentration	[11e,39]
Hydrogenation properties	1) Metal-Metal bond-energy 2) Metal-Hydrogen bond-energy 3) Metal-Metal bond-length 4) Metal-Hydrogen bond-length 5) Volume alteration during hydrogenation 6) Energy alteration during hydrogenation 7) Enthalpy of formation of binary hydrides 8) Pauling electronegativity 9) Proportion of hydrogen affinity elements 10) d-band center 11) HOMO 12) LUMO ...	[11d,e,40]
De-/hydrogenation conditions	1) Atmosphere 2) Heating rate 3) Holding time 4) Temperature 5) Pressure 6) Catalysts 7) Materials to catalysts ratio 8) Flow rate 9) Shape-based descriptions	[11e,41]

essential features from Magpie framework that are already suitable for well-trained ML models. Although down-selection methods offer significant advantages in reducing the feature number and improving model efficiency, they also present potential issues such as the possible loss of useful information, algorithm instability, and sensitivity to data types.<sup>[43,49]</sup>

Dimensionality reduction is a technique for converting data from high-dimensional to low-dimensional representations, aiming at preserving the essential characteristics of the data while reducing its complexity and redundancy, the most frequently used of which are principal component analysis (PCA), non-negative

matrix factorization (NMF), manifold learning algorithm, and so forth.<sup>[50]</sup> PCA determines a series of orthogonal vectors (i.e., principal components) as new features (combinations of the original features), on which the variance of data points is maximized. This means that PCA can identify the primary directions of variation in the data and perform dimensional reduction along these directions, thereby addressing the dimensionality curse that denotes the phenomenon where data becomes sparse and difficult to handle in high-dimensional spaces.<sup>[30a,51]</sup> Currently, PCA has been successfully applied in materials research fields such as supercapacitors, perovskites, composites, and aluminum

alloys.<sup>[52]</sup> However, if the original data contains a significant amount of noise or irrelevant features, these features may also be retained in the principal components by PCA, potentially impacting the performance of the model.<sup>[30,53]</sup> As for NMF, it requires that all matrix elements after the factorization are non-negative, and this constraint makes NMF more interpretable for image processing, text mining, and so on.<sup>[54]</sup> Among the manifold learning algorithms species, t-SNE is the most useful one. The purpose is to find a two-dimensional representation of the data to maintain the distance between the data, so that closer points in the original feature space remain closer. However, t-SNE is mainly used for visualization and therefore rarely for generation of more than two new features.<sup>[43,55]</sup>

### 2.3. Machine Learning Models

The selection and implementation of ML models are paramount for achieving efficient and precise predictions and problem-solving outcomes. Depending on the learning approach, ML can be categorized as supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Among them, supervised learning and unsupervised learning are the most frequently studied in the field of materials and will be highlighted in this section.<sup>[37a,b,56]</sup>

Supervised learning is an ML algorithm based on training sets with explicit target outputs, and can also generate explicit target outputs on data to be tested/validated.<sup>[9e,57]</sup> Based on the continuity of the output results, supervised learning tasks can be classified into regression and classification. The former has continuous numerical outputs, e.g., absorption energy, formation energy, etc., while the latter has predicted discrete-value outputs, e.g., benign versus malignant judgment of breast tumors, classification of specific information, etc.<sup>[9b,58]</sup> General supervised learning encompasses a diverse range of algorithms, such as linear models, support vector machine,  $k$ -nearest neighbors, decision tree, neural network, and ensemble models.<sup>[10a,13c,59]</sup> In contrast, unsupervised learning refers to the situation where there is no clear target output for the inputting data, and the data analogy is unknown, i.e., the computer is expected to learn and output on its own. typical applications include clustering algorithms and dimensionality reduction algorithms (described in Section 2.2.2).<sup>[60]</sup> Next, a brief description of the popular algorithms will be presented.

#### 2.3.1. Linear Models

Linear model, the most fundamental ML algorithm, endeavors to identify the optimal linear relationship that exists between feature values and target values. The general formula can be succinctly expressed as:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p} + c \quad (1)$$

where  $y_i$  is the observed true value of  $i^{\text{th}}$  sample;  $p$  is the total feature number;  $x_{i,1}, x_{i,2}, \dots, x_{i,p}$  are the feature values of  $i^{\text{th}}$  sample;  $\beta_0$  is the intercept term;  $\beta_1, \beta_2, \dots, \beta_p$  are the regression coefficients;  $c$  represents the error term.

Linear models boast advantages such as simplicity, highly interpretability, and straightforward comprehension of results.

Nevertheless, they are prone to outliers and exhibit instability in coefficient estimates due to high correlations among independent variables.<sup>[61]</sup> To mitigate issues of multicollinearity and overfitting, incorporating regularization terms that adeptly select salient features is advisable, as exemplified by the following formulation:

$$\text{loss} = \min_{\beta_0, \beta_1, \dots, \beta_p} \left\{ \sum_{i=1}^n (y_i - \hat{y}_i)^2 + l \right\} \quad (2)$$

where  $\hat{y}_i$  is predicted value, and  $l$  refers to the penalty term.

Based on the type of regularization term, regularization methods can be broadly classified into  $l_1$ ,  $l_2$ , and  $l_1+l_2$ . A typical linear model that employs  $l_1$  regularization is LASSO (Least Absolute Shrinkage and Selection Operator). As expressed in Equation (3),  $l_1$  penalty term considers the sum of the absolute values of the coefficients, and thus the feature of LASSO is its ability to shrink the coefficients of insignificant variables to zero.

$$l_1 = \lambda \sum_{j=1}^p |\beta_j| \quad (3)$$

where  $\lambda$  is the regularization parameter governing the intensity of the penalty term.

In comparison, Ridge introduces the  $l_2$  regularization term and can adjust the coefficients of insignificant variables close to zero. As expressed as Equation (4),  $l_2$  penalty term considers the sum of the squared values of the coefficients.

$$l_2 = \lambda \sum_{j=1}^p \beta_j^2 \quad (4)$$

There are some differences between LASSO and Ridge model to some degree. LASSO tends to produce sparse solutions, which aids in feature selection, but can lead to overfitting or selection of unstable features in some cases.<sup>[62]</sup> Ridge, in contrast, reduces overfitting by shrinking the values of all coefficients, but is less effective at feature selection compared to LASSO.<sup>[63]</sup> Moreover, elastic Net combines the strengths of both LASSO and Ridge (Equation (5)), capable of producing sparse solutions while also reducing overfitting, thereby mitigating some of the limitations that may arise from using either LASSO or Ridge alone.<sup>[64]</sup>

$$l_{\text{elastic-net}} = \lambda_1 \sum_{i=1}^p |\beta_i| + \lambda_2 \sum_{i=1}^p \beta_i^2 \quad (5)$$

Although linear models have advantage of simplicity and strong interpretability to solve simple materials research problems, they still suffer from significant limitations in describing complex nonlinear problems.<sup>[65]</sup>

#### 2.3.2. Support Vector Machine

Draw on kernel functions (linear, polynomial, radial basic function, sigmoid, etc.), support vector machine (SVM) is designed to map original data points into a high-dimensional feature space, enabling it to efficiently handle high-dimensional datasets.<sup>[48b,66]</sup>

SVM encompasses two main categories: support vector classification (SVC) for classification and support vector regression (SVR) for regression.<sup>[67]</sup> However, there are significant differences between SVC and SVR. Specifically, as shown in Figure 3a1, SVC aims to find a hyperplane (two-dimensional space: a straight line; higher dimensions: a hyperplane) that separates sample points of different categories while maximizing the distance from the hyperplane to the fewest and closest sample points (support vectors). These support vectors determine the width of the classification boundaries and the performance of the classifier. In comparison, SVR seeks to find a function (hyperplane or hypersurface) that fits best to the training data, minimizing the distance (i.e., error) from all data points to this function while maintaining a certain margin to prevent overfitting (Figure 3a2). Compared to linear models, SVM exhibits stronger generalization capabilities, and demonstrates robustness against outliers, and is adept at handling nonlinear relationships.<sup>[68]</sup> Several of its applications are also found in the field of materials science. For example, He et al.<sup>[69]</sup> investigated the stability of MXenes two-dimensional materials with a small dataset as well as SVM model. Rahimi et al.<sup>[70]</sup> trained a high-performance SVM model and succeeded in predicting the hydrogen absorption behaviors of activated carbons. Nevertheless, the performance of SVM is heavily dependent on the selection of kernel function and regularization parameters. Furthermore, SVM requires significant memory resources and time when dealing with large-scale datasets.

### 2.3.3. K-Nearest Neighbors

K-nearest neighbors (KNN) is a basic ML technique broadly applicable to classification and regression tasks. The core principle is that, for a given test sample, the algorithm will find the  $k$  closest training samples in the training dataset according to the specific distance metric (e.g., Euclidean distance, Manhattan distance, or Minkowski distance).<sup>[66,71]</sup> Subsequently, the information of these  $k$  “neighbors” will be taken to make predictions. In classification tasks, the majority class among them determines the outcomes, while for regression missions, their average value serves as the results.<sup>[72]</sup> A significant advantage of KNN is its simplicity and intuition, making it not only understandable but also convenient to implement. Unlike other algorithms, KNN merely stores the training dataset during training and makes predictions based on similarity (measured by distance) between test and training samples. However, a potential drawback of KNN is that when the training dataset is very large, the computational cost skyrockets as it requires the computation of the distance between each test and all training samples. In addition, to ensure fairness in distance calculations, preprocessing like normalization or standardization is commonly required to eliminate disparities in feature magnitudes or distributions.<sup>[30a]</sup> Because of its interpretability and simplicity, KNN is also applicable in materials fields as phase analysis, concrete compressive strength, characteristics of piezoelectric ceramics, etc.<sup>[73]</sup>

### 2.3.4. Decision Tree

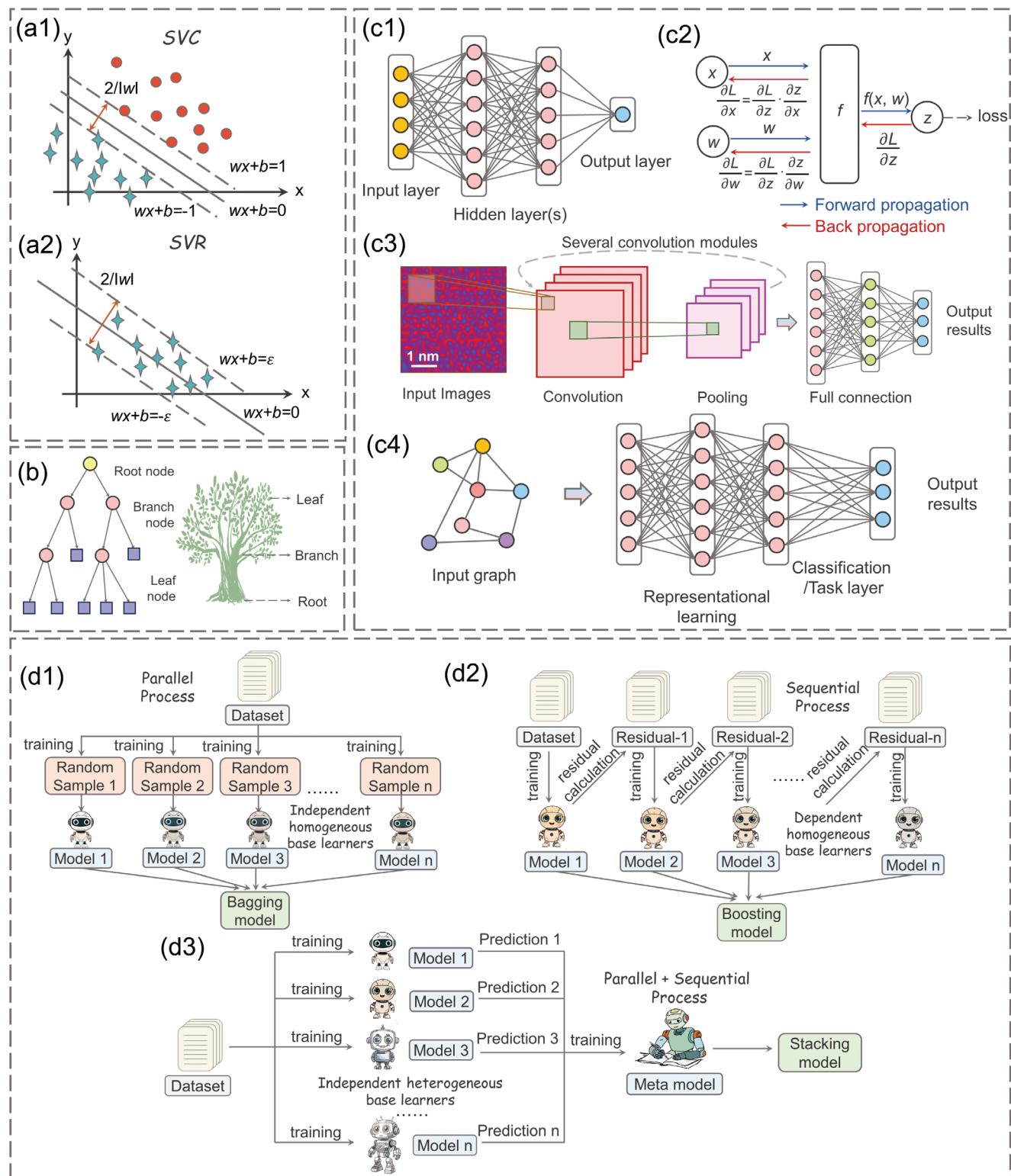
Decision tree (DT) is a classification and regression ML model rooted in a tree-like structure (Figure 3b). It recursively partitions

the dataset into subsets to form a decision tree, where the internal nodes represent judgments on feature attributes, the branches indicate different outcomes of that judgment, and the leaf nodes correspond to the final class labels or predicted values.<sup>[30a,57a,74]</sup> The advantages of DT lie in its fast-training speed, ease of understanding, and the ability to provide reasonable insights through feature importance analysis.<sup>[11c,75]</sup> Nevertheless, the complexity of the model often culminates in overfitting, as it captures noise and outliers from training data rather than genuine patterns. This can be alleviated via pruning, restricting tree depth, and setting a sample threshold for leaf nodes.<sup>[43,76]</sup> Furthermore, the sensitivity of DT to minute training data variations and the neglect of feature correlation during construction may not always yield optimal results. Fortunately, these limitations can be addressed through ensemble techniques like bagging and boosting that will be presented in Section 2.3.6. Despite the drawbacks of the DT model described above, including not being the optimal prediction model in certain situations (weighing the balance between simplicity and accuracy), it remains a fundamental and important model choice that can achieve significant computational efficiency in ML research within the field of materials science.<sup>[77]</sup> For example, Tiwari et al.<sup>[41b]</sup> utilized a simple yet highly precise DT model (compared to XGBoost and RF models) to predict the de-/hydrogenation kinetics.

### 2.3.5. Neural Network

Neural network (NN), inspired by biological neurons, comprises an input layer for receiving data, hidden layers processing weighted inputs from the previous layer through nonlinear activation functions (e.g., Sigmoid, ReLU, Tanh, etc.), and an output layer generating predictions.<sup>[12e,g,56d]</sup> Thus, these networks can learn complex relationships between variables and uncover hidden patterns within the data. In materials research, popular NN models include feedforward neural network (FNN), back propagation neural network (BPNN), convolutional neural network (CNN), and graph neural network (GNN).<sup>[78]</sup>

FNN is a typical structure of an artificial neural network (ANN). As shown in Figure 3c1, within the FNN, information propagates exclusively in a unidirectional manner, traversing from the input layer to the hidden layer and subsequently to the output layer, without any direct feedback loops or recurrent connections throughout the process.<sup>[30a,43,79]</sup> However, the optimization of FNN’s parameters (including weights and biases) typically requires iterative algorithms such as gradient descent, rendering it suitable for simple pattern recognition tasks.<sup>[80]</sup> In contrast, BPNN employs a special error backpropagation algorithm (Figure 3c2). This algorithm calculates the error between the prediction outcomes of the output layer and the desired output results, and then propagates this error back to the hidden layers and even the input layer, adjusting the connection weights between neurons layer by layer.<sup>[81]</sup> Furthermore, since the backpropagation algorithm effectively utilizes gradient information to guide parameter updates, it can generally improve the optimization efficiency of the model to a certain extent. Huang et al.<sup>[82]</sup> successfully predicted the absorption behaviors of atoms on single vacancy graphene surfaces throughout the periodic table using the BPNN model. Besides, BPNN also performs well in predicting



**Figure 3.** Schematic diagram of ML models: a1) SVC, a2) SVR, b) DT, c1) FNN, c2) BPNN, c3) CNN, c4) GNN, d1) bagging, d2) boosting, and d3) stacking models.

perovskite formation energy, thermodynamic stability, and oxygen vacancy formation ability, etc.<sup>[83]</sup>

CNN, a neural network specifically designed for image analysis, consists of multiple convolutional layers, pooling layers, and one or more fully connected layers (Figure 3c3).<sup>[58,84]</sup> These layers work together to extract and recognize complex features within images. Convolutional layers utilize filters to extract local features from the image and generate feature maps. While the Pooling layers play a pivotal role in diminishing the spatial dimensions of feature maps, thereby reducing computational complexity and enhancing the model robustness and generalization capabilities. Finally, fully connected layers map the extracted features to the output space, which can be used for predicting image values or classification labels. CNNs, with their formidable feature extraction and pattern recognition capabilities, have found widespread application in the analysis of material microstructures, DFT-based prediction of surface catalysis/reactivity, etc. For example, Fang et al.<sup>[85]</sup> classify wafer defects by CNN and accurately detected and classified multi-type defects from scanning electron microscope (SEM) images. Back et al.<sup>[86]</sup> successfully determined the binding energy of surface CO and H based on the atomic surface structure-based CNN model.

GNN is a NN model specifically designed to process graph-structured data (nodes, edges, etc.) that represent complex relationships between entities (Figure 3c4). GNN updates node representations by means of information propagation and aggregation among nodes, thereby capturing both global and local characteristics of the graph.<sup>[37b,87]</sup> Compared to traditional NN models, GNN utilizes the connectivity between nodes for information propagation and learning, thus enabling it to better handle complex data structures and relationships. GNN has significantly advanced the field of material property prediction. In the crystal graph convolutional neural network (CGCNN) proposed by Xie et al.,<sup>[37a]</sup> each crystal was represented as a graph, and the invariance of atomic indices and unit cell selection was ensured through convolutional-pooling operations, whereby superior performance in terms of energy-band electronic structure and elasticity properties were achieved. Chen et al.<sup>[88]</sup> developed a materials graph network (MEGNet) model that integrates global state properties, and provided a highly adaptable framework for predicting crystal properties.

NN models typically require a lengthy training process and rely on vast amounts of training data to optimize their performance. However, in specific material domains, acquiring such large quantities of high-quality data is challenging, leading to potential model bias and overfitting.<sup>[89]</sup> Furthermore, the complexity of NN models makes their internal mechanisms difficult to comprehend and interpret. Additionally, these models exhibit a strong dependence on descriptors and require further research and development.

### 2.3.6. Ensemble Models

Although a single ML algorithm can demonstrate superior performance on some issues, ensemble learning strategy, i.e., combining multiple homogeneous or heterogeneous base learners, has emerged as a widespread practice for the sake of reducing the

random fluctuation and error more effectively. Typical ensemble methods that applied in materials science include bagging, boosting, and stacking.<sup>[15a,90]</sup>

As shown in Figure 3d1, bagging is a technique that reduces generalization error by combining multiple homogeneous weak base learners. It generates multiple training subsets through random sampling with replacement from the original dataset, with each subset independently trained to a base learner. Ultimately, the prediction results from these base learners are combined through voting or averaging.<sup>[36,91]</sup> RF, an exemplary representative of the bagging algorithm, extends the concept of bagging by constructing a forest of DTs with each tree trained on random subsets of features and data. This randomness reduces model correlation and improves generalization, making RF robust in supervised learning tasks. For example, Kronberg et al.<sup>[92]</sup> combined the RF model to study the hydrogen absorption properties and key influencing factors of defective nitrogen-doped carbon materials. Similarly, Kusdhany et al.<sup>[11b]</sup> and Thanh et al.<sup>[93]</sup> investigated the manner in which temperature, chemical properties, and structural features affect the hydrogen storage capacity of porous carbon materials with the support of RF models.

Bagging reduces variance by building independent homogeneous weak learners in parallel and is suitable for high-stability prediction scenarios. Differently, boosting reduces bias by building dependent homogeneous weak learners sequentially and is suitable for high-accuracy prediction scenarios. Depending on the optimization objective, boosting methods can be classified into AdaBoost and gradient boosting. The former optimizes the model by adjusting the weights of the samples that are mispredicted by the current learner, and the latter optimizes the model by minimizing the residuals of the previous learner's model (Figure 3d2).<sup>[94]</sup> Typical ML models that employ this technique include gradient boosting decision tree (GBDT) and XGBoost (regularization considered).<sup>[30a,95]</sup> For instance, Chen et al.<sup>[96]</sup> implemented a GBDT model to guide the optimization of the structural parameters and the operational parameters during the control process of the hydrogen fuel cells. Yang et al.<sup>[97]</sup> adopted a well-trained XGBoost model to accurately describe the electrochemical behaviors of biochar. Generally, although boosting algorithms may require longer training time compared to bagging due to their sequential construction process, this approach often leads to relatively higher prediction accuracy.<sup>[30a,98]</sup>

The base learner for bagging and boosting is typically DT, whereas stacking is constructed based on multiple heterogeneous base learners and utilizes a meta-model to achieve the final output.<sup>[99]</sup> As displayed in Figure 3d3, the acquisition of the stacking model comprises two stages. In the first stage, multiple heterogeneous base models are adopted to make predictions on the training data, and these predictions are then used as new features to input into the second stage. In the second stage, a meta-model is employed to make predictions based on the outputs from the first stage. Although the training process of the stacking model is relatively complicated and time-consuming, its high effectiveness and accuracy in practice make it a preferred method for materials research.<sup>[43,57a]</sup> For instance, Wang et al.<sup>[100]</sup> established a stacking model using a high-performance baseline model and was able to accurately predict the bandgap of

novel semiconductor materials. Li et al.<sup>[101]</sup> constructed a 2D-3D stacking model based on 3D atomic positions-based DimeNet++ and 2D connectivity graph-based GCN, which can deal with fast and accurate scaling of quantum chemical computational data.

By incorporating the multiple prediction models, ensemble learning can significantly enhance prediction accuracy, generalization ability, training efficiency, and model robustness and reliability. Aside from that, it can effectively reduce bias and variance, mitigate overfitting risks, handle complex data, accelerate training pace, and provide multi-task applicability and cross-domain flexibility.<sup>[14b,31a,41a,102]</sup>

### 2.3.7. Clustering

Clustering is a category of unsupervised learning algorithms that efficiently uncovers the potential patterns in unknown data by partitioning unlabeled data into groups, each containing similar objects. Common clustering algorithms applied in materials research include *k*-means clustering, hierarchical clustering, DBSCAN (density-based spatial clustering of applications with noise), etc.<sup>[12e,30a,43,103]</sup>

As one of the most representative clustering methods, *k*-means clustering assigns points in a dataset to *k* clusters through an iterative process, minimizing the sum of the distances between each point and the centroid (i.e., the mean of all points within the cluster) of its assigned cluster.<sup>[104]</sup> The value of *k* is preset by researchers based on prior knowledge or experimental requirements. This method is computationally efficient when handling continuous *n*-dimensional spatial data, but its performance is notably affected by the initial selection of cluster centers and the presence of noisy data, and it can only detect/generate spherical clusters.<sup>[30a]</sup> As for the hierarchical clustering, it is known for building a tree-like hierarchical structure of data points. The algorithm can be either agglomerative, which merges smaller clusters into larger ones from bottom to top, or divisive, which splits larger clusters into smaller ones from top to bottom. Hierarchical clustering does not require specifying the number of clusters in advance, providing flexibility in cluster selection. While the interpretation of results and the determination of final clusters generally depend on the subjective judgment of researchers.<sup>[105]</sup> Additionally, this algorithm may incur higher computational costs when handling large-scale datasets and is also sensitive to noise. DBSCAN overcomes the sensitivity of *k*-means and hierarchical clustering to noise and outliers, and can identify clusters of arbitrary shapes. DBSCAN identifies high-density regions as clusters and low-density regions as outliers by defining two parameters: neighborhood size and minimum number of samples. This algorithm is particularly suitable for datasets with complex shapes and varying densities, but the computational complexity is high and the choice of parameters may significantly affect the clustering results.<sup>[27c,106]</sup>

### 2.3.8. Machine Learning Interatomic Potentials

ML models presented in Sections 2.3.1–2.3.7 are primarily tailored for predicting material performance based on various vari-

ables or features, including elemental composition, physicochemical properties, and structure. However, these models lack a profound understanding of the mechanisms underlying atomic-level interactions.

Machine learning interatomic potentials (MLIPs) are high-precision interatomic interaction potentials developed using ML techniques. By training on extensive datasets generated through high-precision first-principles calculations, MLIPs can accurately capture and describe interatomic interactions.<sup>[107]</sup> Compared to empirical interatomic potentials, MLIPs demonstrate superior accuracy. When compared to DFT, MLIPs significantly boost computational efficiency.<sup>[108]</sup> Thus, the development of MLIPs with both high accuracy and efficiency poses a significant research challenge for the future. Similarly, the workflow for creating MLIPs involves dataset construction, model training, and model validation. A precise description of the crystal structure is crucial for constructing accurate MLIPs. Due to the unique symmetry properties of crystal structures, such as translational and rotational symmetry, it is not feasible to use Cartesian coordinates directly as input. Instead, descriptors based on coarse coordinates, like interatomic distances and angles, are used to ensure that the symmetry of the system is preserved. By combining these descriptors, researchers have developed special neural network architectures, such as high-dimensional neural network potentials<sup>[109]</sup> based on atom central symmetry functions (ACSF)<sup>[110]</sup> and deep potential energy models.<sup>[111]</sup> These models can better capture the symmetry features of crystal structures, thereby improving prediction accuracy. Another method involves adopting equivariant neural network potential function models, in which transformations of the input data are reflected in the network's output. For example, Nequ IP is an equivariant neural network potential function model based on the E(3) group, which can achieve prediction accuracy comparable to that of DFT even with smaller datasets.<sup>[112]</sup>

MLIPs allow for the conduct of large-scale MD simulations with an accuracy that is nearly comparable to first-principles calculations. This capability allows researchers to investigate material behavior and explore novel structures with unprecedented detail and precision. For example, Guo et al.<sup>[113]</sup> successfully explored novel SnSe structures with quotient graph and deep learning potential. Dai et al.<sup>[114]</sup> utilized the Deep Potential Molecular Dynamics (DeePMD) MLIP to predict the elastic properties of the high-entropy (Zr<sub>0.2</sub>Hf<sub>0.2</sub>Ti<sub>0.2</sub>Nb<sub>0.2</sub>Ta<sub>0.2</sub>)C, and the predicted values align well with the experimental measurements. Regarding the application of MLIPs in solid-state hydrogen storage, the focus is on Mg-based HSMs, and a comprehensive overview of the pertinent research will be presented in Section 3.5.

### 2.4. Model Evaluation Metrics

Through algorithm selection and hyperparameter tuning, it is possible to construct multiple ML models tailored to a specific task. For different common ML tasks (supervised regression/classification, and unsupervised tasks), it is essential to employ different evaluation metrics to comprehensively and accurately assess model performance.<sup>[12b,13b,115]</sup>

As displayed in Equations (6–9), the main evaluation metrics for supervised regression missions include coefficient of

determination ( $R^2$ ), mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), etc.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \quad (7)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

where  $\bar{y}$  is the average observed value of total n samples. A regression model with an  $R^2$  value closer to 1 and smaller MAE, MSE, and RMSE values would indicate a more accurate model.

When it comes to supervised classification mission, precision, recall, accuracy, Confusion Matrix, F1 score, receiver operating characteristic (ROC) curve, area under curve (AUC), etc. are the more suitable choices (some are expressed in Equations (10–13)).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (12)$$

$$F1 \text{ score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

where TP denotes correctly labeled positive cases (true positive), TN denotes correctly labeled negative cases (true negative), FP denotes incorrectly labeled positive cases (false positive) and FN denotes incorrectly labeled negative cases (false negative). TP, FP, TN, and FN can all be obtained from a confusion matrix. Specifically, precision represents the model's confidence in positive class, while recall reflects the model's ability to capture positive samples. Meanwhile, accuracy is the most straightforward performance evaluation metric, but it may not be accurate enough in cases of class imbalance. Besides, a higher F1-score indicates better model performance, as it incorporates both precision and recall, avoiding potential misleading interpretations from a single metric. ROC curve is derived by varying the classification threshold with the false positive rate and the true positive rate as the horizontal and vertical axes, respectively. The closer the ROC curve approaches the upper left corner, indicating an AUC value nearing 1, signifies superior classification performance.

In the context of unsupervised learning, it generally consists of adjusted rand index (ARI), normalized mutual information (NMI), silhouette coefficient (SI), etc. The value range of ARI

range spans from -1 to 1, yet it typically resides within 0 to 1. Higher value indicates greater similarity between clustering outcomes. Likewise, NMI also varies between 0 to 1, where higher value reflects a greater degree of shared information between clustering results and superior clustering performance. As for SI, it ranges from -1 to 1. A value close to 1 indicates that the samples are well-clustered, while a value close to -1 suggests the samples should be reassigned.

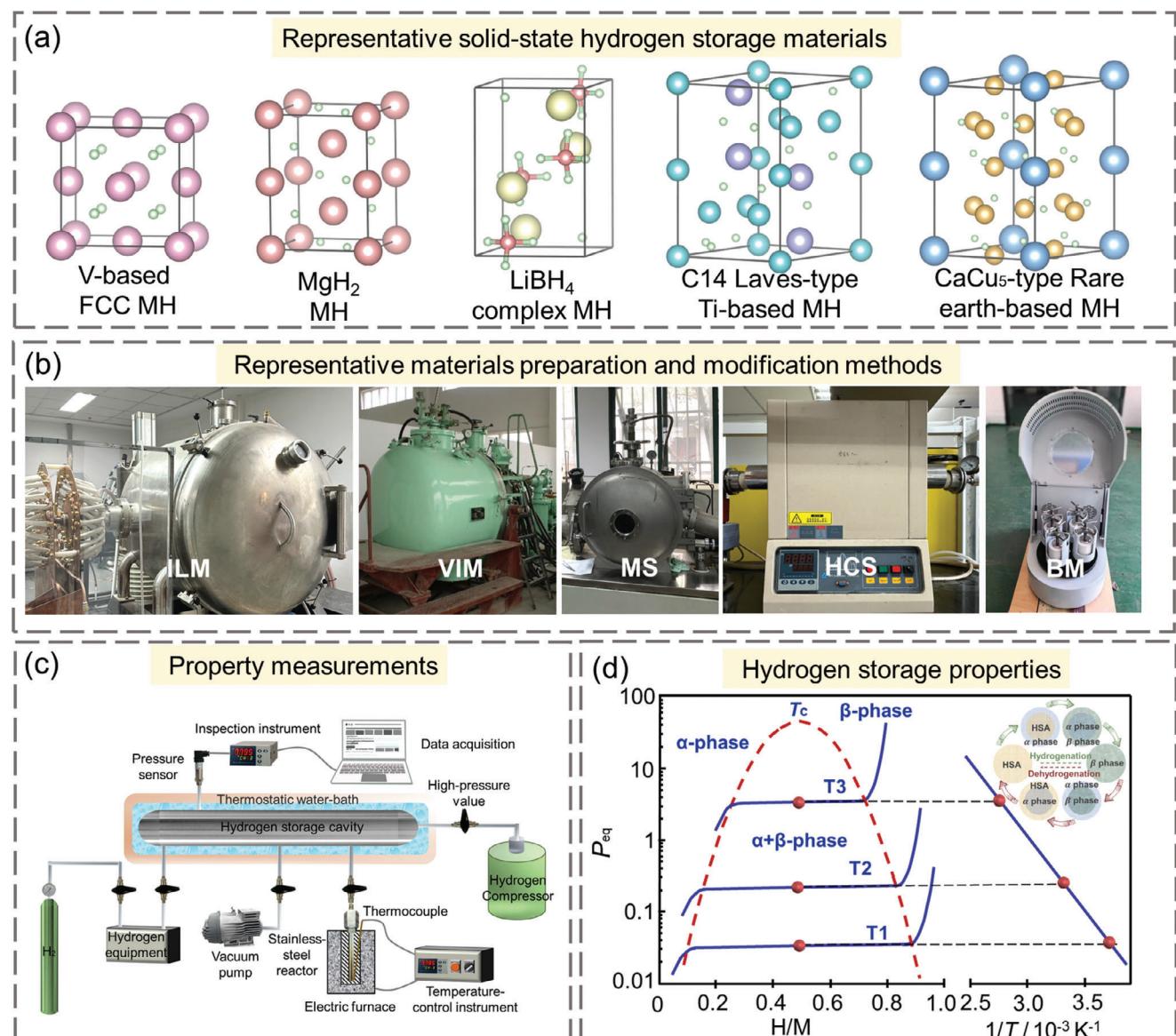
### 3. Applications of Machine Learning for Solid-State Hydrogen Storage

A variety of solid-state HSMs have been developed, including titanium-based HSMs, rare-earth-based HSMs, solid solution HSMs, Mg-based HSMs, and complex HSMs, which can be synthesized or modified by techniques such as induction levitation melting (ILM), vacuum induction melting (VIM), melt spinning (MS), hydriding combustion synthesis (HCS), and ball milling (BM), as illustrated in Figure 4.<sup>[2a,39b,116]</sup> As an intrinsic character of solid-state HSMs, reversible de-/hydrogenation is essentially a process in which atomic hydrogen preferential occupation/de-embedding in/from the tetrahedral/octahedral interstitial sites (metal hydrides) or forms/breaks stable chemical bonds (Mg-based and complex HSMs). This process involves a thermodynamic and kinetic mass and heat transfer equilibrium, thus reflecting different hydrogen storage performance, such as hydrogen storage capacity, de-/hydrogenation equilibrium pressure, pressure-composition-temperature (PCT) plateau characteristics, thermodynamic enthalpy/entropy, kinetics, mass and heat transfer efficiency, etc.<sup>[117]</sup> In the following part, the general and specialized patterns of solid-state HSMs that have been explored by ML will be presented and discussed, respectively.

#### 3.1. General Patterns of Solid-State HSMs

As introduced in Section 2.1, HydPARK dataset comprises a diverse range of solid-state HSMs and their de-/hydrogenation performance, enabling ML model to provide general insight into the intricate composition-structure-property relationships. Currently, ML is being actively used to explore various facets of solid-state HSMs with HydPARK dataset, including hydrogen storage capacity, hydrogen storage systems classification, formation energy, hydride formation enthalpy, equilibrium pressure, novel material discovery, etc., as displayed in Table 3.

Hydrogen storage capacity is the primary characteristic of all HSMs. Rahnama et al.<sup>[27a]</sup> demonstrated that GBDT model is the most accurate in describing the hydrogen storage capacity with  $R^2$  of 0.83. In addition, it was revealed that material category, temperature, and hydride formation enthalpy are paramount factors influencing hydrogen storage capacity, as evidenced by their high feature importance ranking. Conversely, the material's specific composition has a relatively negligible impact on the amount of hydrogen storage. Given the substantial variations in hydrogen storage behavior mechanisms and saturated capacity among different material systems, the above conclusions hold true for the diverse categories of HSMs encompassed within the HydPARK



**Figure 4.** a) Some commonly available metal hydrides (MH); b) Typical preparation methods for HSMs: ILM, VIM, MS, HCS, BM, etc.; c) Sievert-type apparatus for the measurements of hydrogen storage performance; d) PCT curves and Van't Hoff plot.

dataset, whereas they may not be applicable to a particular material system. To effectively differentiate between various HSM categories, Rahnama et al.<sup>[27b]</sup> achieved an accuracy of 0.80 utilizing multiclass NN classifier. Besides, the observed positive correlation among gravimetric hydrogen storage capacity, hydride formation enthalpy, and temperature indicates that complex hydrides possess the potential to achieve the highest gravimetric hydrogen storage capacity at elevated temperatures, with a broad operating temperature range. Beyond the supervised learning task mentioned, Rahnama et al.<sup>[27d]</sup> employed the *k*-means unsupervised clustering model to explore the similarities among various material categories in terms of their gravimetric hydrogen storage capacity and experimental parameters. This cluster analysis reveals that samples with identical structures could be segregated into three distinct clusters, emphasizing a notable correlation be-

tween A<sub>2</sub>B-type intermetallic compounds, complex hydrides, and Mg-based materials with respect to temperature, pressure, and hydrogen storage capacity.

The prediction of formation energy facilitates the high-throughput screening of the thermodynamic feasibility of arbitrary novel hydride formulations and mixtures. Drawing on the HydPARK dataset and the MP database, Nations et al.<sup>[11b]</sup> demonstrated the predictive capability of simple compositional attributes in estimating formation energy, where the RF ensemble model yielding the optimal performance (MSE: 0.102 eV<sup>2</sup>). Remarkably, both well-explored formulations involving alkali and alkaline earth metals, and less-explored compositions incorporating transition metals and lanthanides can perform well doped with boron. Ultimately, the authors further created a web-based application that employed this model to simplify the

**Table 3.** Machine learning research solid-state hydrogen storage materials.

Material class	Dataset used and refs	Feature descriptors	Target properties	Machine learning models	Best model	Evaluation metrics	Refs.
Multi-category HSMs	HydPARK <sup>[25]</sup>	1) Materials class 2) Temperature 3) Heat of formation 4) Pressure 5) Composition formula 1) Hydrogen storage capacity 2) Temperature 3) Heat of formation 4) Pressure 5) Hydrogen storage capacity 1) Materials class 2) DF 3) DJ 4) NN	Hydrogen storage capacity	1) LR 2) NN 3) BLR 4) GBDT	GBDT	MAE: 0.003 wt% H <sub>2</sub> RMSE: 0.012 wt% H <sub>2</sub> $R^2$ : 0.83	[27a]
Multi-category HSMs	HydPARK <sup>[25]</sup>	1) Temperature 2) Pressure 3) Heat of formation 4) Pressure 5) Hydrogen storage capacity 1) Materials class 2) DF 3) DJ 4) NN	Material class	1) LR 2) DF 3) DJ 4) NN	NN	Accuracy: 0.8	[27b]
Multi-category HSMs	HydPARK <sup>[25]</sup>	1) Hydrogen storage capacity 2) Temperature 3) Heat of formation 4) Pressure 5) Hydrogen storage capacity 1) Temperature 2) Pressure 3) Heat of formation 4) Pressure 5) Hydrogen storage capacity 1) Materials class 2) DF 3) DJ 4) NN	Similar behaviors of different materials classes	k-means	Number of clusters: 3 Intra-cluster cohesion: 0.6 Inter-cluster separation: 1.4	[27d]	
Multi-category HSMs	Materials project <sup>[17]</sup> + HydPARK <sup>[25]</sup>	1) Band gap 2) Density 3) Atomic volume 4) Magnetization 5) Electronegativity 6) Average <i>d</i> -character 7) Average <i>f</i> -character 8) Temperature 9) Pressure	Formation energy	1) Ridge 2) DT 3) RF 4) GBDT	RF	MSE: 0.102 eV <sup>2</sup>	[118]
Multi-category HSMs	HydPARK <sup>[25]</sup>	1) Fundamental properties of elements 2) Interactions-based features 3) Composition-dependent properties 4) Temperature	Equilibrium pressure ( $\ln P_{eq}^0$ ) Hydride formation enthalpy	1) RepTree 2) RF 3) NN	GBDT	MAE <sub>train</sub> = 0.47 MAE <sub>test</sub> = 1.52	[27c]
Multi-category HSMs	HydPARK <sup>[25]</sup>	Magpie framework	Equilibrium pressure ( $\ln P_{eq}^0$ ) Hydride formation enthalpy	1) RepTree 2) RF 3) NN	RF	MAE: 8.56 kJ mol <sup>-1</sup> MRE: 28%	[27f]
Multi-category HSMs	HydPARK <sup>[25]</sup>	1) Hydrogen storage capacity (H-ST) 2) Hydride formation enthalpy (THOR)	1) Hydrogen storage capacity (H-ST) 2) Hydride formation enthalpy (THOR)	1) ETR 2) ETR	1) ETR 2) ETR	$1. R^2 = 0.81$ MAE = 0.45 wt% H <sub>2</sub> $2. R^2 = 0.89$ MAE = 4.53 kJ mol <sup>-1</sup> H <sub>2</sub>	[11e]
		3) KRR 4) GPR 5) RF 6) RF 7) ETR 8) GBDT					

(Continued)

**Table 3.** (Continued)

Material class	Dataset used and refs	Feature descriptors	Target properties	Machine learning models	Best model	Evaluation metrics	Refs.
Ti-based HSMs (AB <sub>2</sub> -type HSMs)	Proprietary dataset <sup>[32]</sup>	Constituent elements	1) Hydride formation enthalpy 2) Phase abundance 3) Hydrogen storage capacity	1) MLR 2) DT 3) RF	1) RF 2) RF 3) RF	1. $R^2 = 0.647$ MAE = 4.36 kJ mol <sup>-1</sup> H <sub>2</sub> RMSE = 5.75 kJ mol <sup>-1</sup> H <sub>2</sub> 2. $R^2 = 0.832$ MAE = 22.4% RMSE = 25.6% 3. $R^2 = 0.688$ MAE = 0.101 wt% H <sub>2</sub> RMSE = 0.159 wt% H <sub>2</sub>	[32]
Ti-based HSMs (AB <sub>2</sub> -type HSMs)	Proprietary dataset <sup>[32]</sup>	Constituent elements	Hydride formation enthalpy	GPR with different kernel	GPR with Exponential kernel	$R^2 = 0.969$ MRE = 2.291% MSE = 3.909 kJ <sup>2</sup> mol <sup>-2</sup> H <sub>2</sub> RMSE = 2.501 kJ mol <sup>-1</sup> H <sub>2</sub> STD = 1.878 kJ mol <sup>-1</sup> H <sub>2</sub>	[120]
Ti-based HSMs (AB <sub>2</sub> -type HSMs)	Proprietary dataset <sup>[32]</sup>	Constituent elements	Hydrogen storage capacity	1) 1.GA-LSSVM 2) 2.PSO-LSSVM 3) 3.HGAPSO-LSSVM	HGAPSO-LSSVM	$R^2 = 0.980$ STD = 0.043 wt% H <sub>2</sub> MSE = 0.002 wt% <sup>2</sup> H <sub>2</sub> RMSE = 0.045 wt% H <sub>2</sub>	[121]
Ti-based HSMs (AB <sub>2</sub> -type HSMs)	Proprietary dataset <sup>[33]</sup>	Constituent elements	Hydride formation enthalpy	1) MLR 2) SVM 3) DT 4) GPR with different kernels	GPR with Exponential kernel	$R^2 = 0.98$ MRE = 2.291% MSE = 1.18 kJ <sup>2</sup> mol <sup>-2</sup> H <sub>2</sub> RMSE = 1.05 kJ mol <sup>-1</sup> H <sub>2</sub> MAE = 0.70 kJ mol <sup>-1</sup> H <sub>2</sub>	[33]
Ti-based HSMs (C14 Laves-type HSMs)	Proprietary dataset <sup>[11]</sup>	1) Magpie framework 2) Constituent elements 3) Atomic radius 4) Hydrogen affinity ability 5) Elemental stoichiometry ratio	1) Cell Volume 2) De/hydrogenation equilibrium pressure ( $\ln \rho_{\text{abs/des}}$ ) 3) Hydride decomposition enthalpy 4) Hydrogen storage capacity	1) LR 2) PLR 3) Ridge 4) LASSO 5) SGDRT 6) DT 7) RF 8) GBDT 9) SVR 10) MLP	1) SVR 2) SVR 3) GBDT 4) GBDT	1) $R^2 = 0.987$ 2) Abs; $R^2 = 0.973$ Des; $R^2 = 0.978$ 3) $R^2 = 0.916$ 4) $R^2 = 0.822$	[11]
Ti-based HSMs (AB <sub>2</sub> -type HSMs)	Proprietary dataset <sup>[16]</sup> (Fitted PCT data)	1) Constituent elements 2) Temperature 3) Pressure	De-/hydrogenation PCT behaviors	1) RF 2) KNN 3) DNN	DNN	$R^2 = 0.9307$	[122]
RE-based HSMs (AB <sub>5</sub> -type HSMs)	Proprietary dataset <sup>[24]</sup>	Fundamental properties of elements	1. Hydride decomposition enthalpy 2. Hydrogen storage capacity	Stepwise regression	1. $YN = 51.2$ $R = 0.99$ 2. $YN = 0.69$ $R = 0.88$	Stepwise regression	[124]

*(Continued)*

Table 3. (Continued)

Material class	Dataset used and refs	Feature descriptors	Target properties	Machine learning models	Best model	Evaluation metrics	Refs.
RE-based HMs ( $\text{AB}_3$ -type HMs)	Proprietary dataset[25]	Constituent elements	1. Maximum discharge capacity 2. Fast discharge performance	SVR-PSO	SVR-PSO	1. MAPE = 2.35% RMSE = 9.74 mAh/g  2. MAPE = 0.89% RMSE = 1.38%  R = 0.91	[25]
Solid solution HMs (V-Ti-Cr-Fe-based HMs)	Proprietary dataset[27]	1) Constituent elements 2) Fundamental properties of elements 3) Temperature	Hydrogen storage capacity	1) LR 2) LASSO 3) XGBoost 4) BF 5) SVM 6) Stacking 7) (RF, SVM, XGBoost) 8) Adaboost SVM 9) Adaboost RF 10) Adaboost XGBoost 11) Adaboost Stacking	Adaboost Stacking	MSE: 0.187 wt% <sup>2</sup> H <sub>2</sub> MAE: 0.307 wt% H <sub>2</sub>  $R^2$ = 0.783 PC = 0.885	[127]
Solid solution HMs	Proprietary dataset[25]	Constituent elements	Hydrogen storage capacity	SVR	SVR	MAPE = 0.18% RMSE = 0.012 wt%  R = 0.99	[125]
Solid solution HMs (V-Ti-Cr-Fe-based HMs)	Proprietary dataset[38b]	1) Constituent elements 2) Fundamental properties of elements 3) Temperature	Hydride decomposition enthalpy	1) LR 2) SVM 3) BF 4) XGBoost 5) Stacking GBDT	XGBoost	MAE: 8.58 kJ mol <sup>-1</sup> RMSE: 11.73 kJ mol <sup>-1</sup>  $R^2$ = 0.783 PC = 0.885	[38b]
Solid solution HMs (High entropy alloys)	ML-HydPARK v0.0.6 (Updated with data for (TiNb)X)[27c,28]	1) Magpie framework 2) Hydride specific feature	1) Hydrogen storage capacity (H/M) 2) Hydride decomposition enthalpy 3) Hydride decomposition entropy	1) Magpie framework 2) Binary hydride formation enthalpy 4) Equilibrium Pressure ( $\ln P_{\text{eq}}^0 / \rho_0$ )	GBDT	1. MAE = 0.12 2. MAE = 4.2 kJ mol <sup>-1</sup> H <sub>2</sub> 3. MAE = 11 J mol <sup>-1</sup> /K H <sub>2</sub> 4. MAE = 1.1  $R^2$ = 0.69	[128]
Solid solution HMs (High entropy alloys)	HydPARK+ (ML-HydPARK v0.1)[27c,28] + two more data)	1) Magpie framework 2) Binary hydride formation enthalpy	Structure-based framework (Extracted by GNN)	GBDT	GBDT	1. MAE: 1.4 2. MAE: 5.5 kJ mol <sup>-1</sup> H <sub>2</sub>	[28]
Solid solution HMs (High entropy alloys)	Proprietary dataset[29]	(Geometrical structure)	Formation energy	GNN	GNN	MAE = 3.1 meV/atom	[129]

(Continued)

**Table 3.** (Continued)

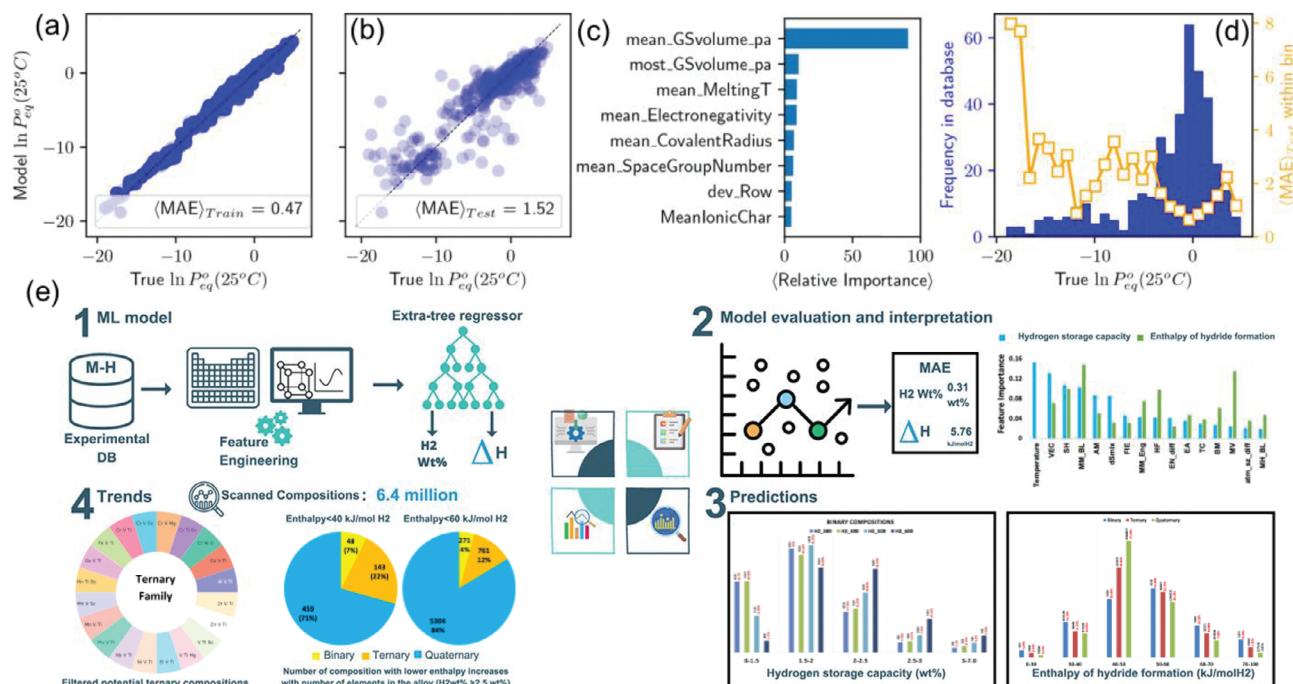
Material class	Dataset used and refs	Feature descriptors	Target properties	Machine learning models	Best model	Evaluation metrics	Refs.
Solid solution HSMs (High entropy alloys)	Proprietary dataset[40b] (DFT dataset)	Fundamental properties of elements	Hydrogen binding energies for the first ( $\Delta E_{\text{mono}}$ ) and second ( $\Delta E_{\text{di}}$ ) stages of absorption	1) LR 2) DT 3) GPR	1.GPR 2. GPR	1. $\Delta E_{\text{mono}}$ : MAE = 0.04 eV R <sup>2</sup> = 0.71 2. $\Delta E_{\text{di}}$ : MAE = 0.04 eV R <sup>2</sup> = 0.93 R > 0.95	[40b]
Mg-based HSMs (Mg-M-H)	Proprietary dataset[132]	Constituent elements	1) Hydrogen storage capacity 2) Dehydrogenation temperature	ANN	ANN		[132]
Mg-based HSMs (Mg-M-H)	Proprietary dataset[133] (DFT dataset)	1) Stoichiometric properties 2) Elemental intrinsic characteristics 3) Atomic number 4) Elemental thermodynamic characteristics 5) Hydrogen storage capacity	3) Plateau pressure Dehydrogenation temperature	1) GBDT 2) RF 3) XGBoost 4) DT	GBDT	R <sup>2</sup> = 0.95 RMSE = 29.98 K	[133]
Mg-based HSMs (Mg-Ni-based HSMs)	Proprietary dataset[134]		Discharge capacities in the initial five cycles	ANN	ANN	R <sup>2</sup> > 0.99	[134]
Mg-based HSMs	Proprietary dataset[34] (DFT dataset)	1) Evaid sum matrices 2) Sine matrices 3) System and frames	Formation energy	1) KRR 2) DeepPMD	DeepPMD	RMSE = 0.43 meV/atom	[34]
Mg-based HSMs (MgH <sub>x</sub> nanoclusters)	Proprietary dataset[135] (DFT dataset)	Structure-based framework	1) Energy from DFT 2) Force from DFT	ANN	ANN	1. RMSE: 31.25 meV/atom MAE: 18.23 meV/atom 2. RMSE: 189.9 meV Å <sup>-1</sup> MAE = 95.48 meV Å <sup>-1</sup>	[135]
Mg-based HSMs (MgH <sub>2</sub> nanoclusters)	Proprietary dataset[137] (DFT dataset)	Structure-based framework	1) Energy of nanoparticles 2) Forces of nanoparticles	SchNet-SSCHA	SchNet-SSCHA	1.MAE: 0.02 eV 2. MAE = 0.02 eV Å <sup>-1</sup>	[137]
Mg-based HSMs (2D MgH <sub>2</sub> sheets)	Proprietary dataset[138] (Material Clouds 2D structure database)[15]	Point-cloud-like image representation of a 2D matrix	Reliability of generated structures	GAN	GAN	Structure validation ratio: 96.8% Structure generation ratio: 87.3%	[138]

(Continued)

Table 3. (Continued)

Material class	Dataset used and refs	Feature descriptors	Target properties	Machine learning models	Best model	Evaluation metrics	Refs.
Mg-based HSMs	Proprietary dataset <sup>[139]</sup>	1) Fundamental properties of elements 2) VEC distributions 3) Thermodynamic parameters 4) Processing conditions	1) Maximum absorption capacity ( $A_{\text{b\_max}}$ ) 2) Maximum desorption capacity ( $D_{\text{e\_max}}$ )	1) SVR 2) MLP 3) GBDT 4) RF	1) GBDT 2) MLP	1. MSE: 0.12 wt% <sup>2</sup> H <sub>2</sub> MAE: 0.24 wt% H <sub>2</sub> $R^2 = 0.947$ 2. MSE: 0.16 wt% <sup>2</sup> H <sub>2</sub> MAE: 0.32 wt% H <sub>2</sub> $R^2 = 0.922$ MAE = 5.4 kJ mol <sup>-1</sup> H <sub>2</sub>	[139]
Mg-based HSMs (Mg-Ni-M ternary intermetallic)	Proprietary dataset <sup>[140]</sup> (Materials project) <sup>[17]</sup>	Structure-based framework	Hydride formation enthalpy	MEGNet+ Transfer learning	MEGNet+ Transfer learning	1. MAE: 13 J mol <sup>-1</sup> /K H <sub>2</sub> $R^2 = 0.87$ 2. MAE: 5.4 kJ mol <sup>-1</sup> H <sub>2</sub> $R^2 = 0.87$ 3. MAE: 13 J mol <sup>-1</sup> /K H <sub>2</sub> $R^2 = 0.67$ 4. MAE: 1.5 J mol <sup>-1</sup> /K H <sub>2</sub> $R^2 = 0.89$	[140]
Mg-based HSMs	ML-HydPARK v0.0 <sup>[141]</sup>	Magpie framework	1) Hydrogen storage capacity (H/M) 2) Hydride formation enthalpy 3) Hydride formation entropy 4) Equilibrium pressure ( $\ln P_{\text{eq}}^0 / P^0$ )	GBDT	GBDT	1. MAE: 0.14 $R^2 = 0.80$ 2. MAE: 5.4 kJ mol <sup>-1</sup> H <sub>2</sub> $R^2 = 0.87$ 3. MAE: 13 J mol <sup>-1</sup> /K H <sub>2</sub> $R^2 = 0.67$ 4. MAE: 1.5 J mol <sup>-1</sup> /K H <sub>2</sub> $R^2 = 0.89$	[30b]
Complex HSMs (LiBH <sub>4</sub> -based mixture)	Proprietary dataset <sup>[41a]</sup>	1) Catalysts and their ratio 2) Ball milling condition 3) Heating rate	Dehydrogenation kinetic performance	1) LR 2) Ridge 3) LASSO 4) MLP 5) GBDT 6) DT 7) Ada DT 8) Bagging 9) ETR 10) RF 11) Hist 12) Ensemble of an ensemble (EOE)	EOF (based on DT&GBDT& KNN&MLP)	MSE = 1.144 wt% <sup>2</sup> H <sub>2</sub> RMSE = 1.066 wt% H <sub>2</sub> EV = 0.889 $R^2 = 0.888$ PC = 0.944 SC = 0.949	[41a]
Complex HSMs (Metal Borohydride)	Proprietary dataset <sup>[144]</sup> (Materials project) <sup>[17]</sup>	Structure-based framework	Formation energy	CGCNN	CGCNN	MAE = 0.063 eV/atom	[144]

Notes (related abbreviations): Machine learning models: Linear Regression (LR), Bayesian Linear Regression (BLR), Multivariate Linear Regression (MLR), Logistic Regression (L-R), Kernel Ridge Regression (KRR), Least Absolute Shrinkage and Selection Operator (LASSO), Polynomial Linear Regression (PLR), SGD Regression (SGDR), Decision Tree (DT), Random Forest (RF), Gradient Boosted Decision Tree (GBDT), Histogram-based Gradient Boosting Tree (Hbst), Decision Forest (DF), Decision Jungle (DJ), ExtraTree Regression (ETR), LGBM Regression (LGBMR), Support Vector Machine (SVM), Least Square Support Vector Machine (LSSVM), Support Vector Regression (SVR), Neural Network (NN), Artificial Neural Network (ANN), MultiLayer Perceptron (MLP), Back Propagation Neural Network (BPNN), Graph Neural Networks (GNN), Crystal Graph Convolutional Neural Network (CGCNN), Generative Adversarial Network (GAN), Materials Graph Network (MEGNet), Deep Potential Molecular Dynamics (DeepMD), Stochastic Self-Consistent Harmonic Approximation (SSCHA), Gaussian Process Regression (GPR), Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Hybrid GA-PCO (HGAPSO), Evaluation metrics: Mean Squared Error (MSE), Mean Relative Error (MRE), Mean Absolute Error (MAE), Rooted Mean Squared Error (RMSE), Correlation Coefficient (R), Coefficient of Determination ( $R^2$ ), Standard Deviation (STD), Residual Standard Error (N), Mean Absolute Percentage Error (MAPE), Pearson Correlation (PC), Spearman Correlation (SC), Explained Variance (EV)



**Figure 5.** a) Training and b) testing results of GBDT model on  $\ln P_{eq}^o$ ; c) eight most important features for  $\ln P_{eq}^o$ ; d) Frequency and  $\langle \text{MAE} \rangle_{\text{test}}$  within each individual bin for the true  $\ln P_{eq}^o$ .<sup>[27c]</sup> a-d) Reproduced with permission.<sup>[27c]</sup> Copyright 2020, American Chemical Society. e) Decoding pathways for solid state hydrogen storage via ML.<sup>[11e]</sup> e) Reproduced with permission.<sup>[11e]</sup> Copyright 2023, Elsevier Ltd.

exploration of existing metal hydride compositions and predict formation energy based on user-defined properties, thereby accelerating the prototyping process for innovative formulations.

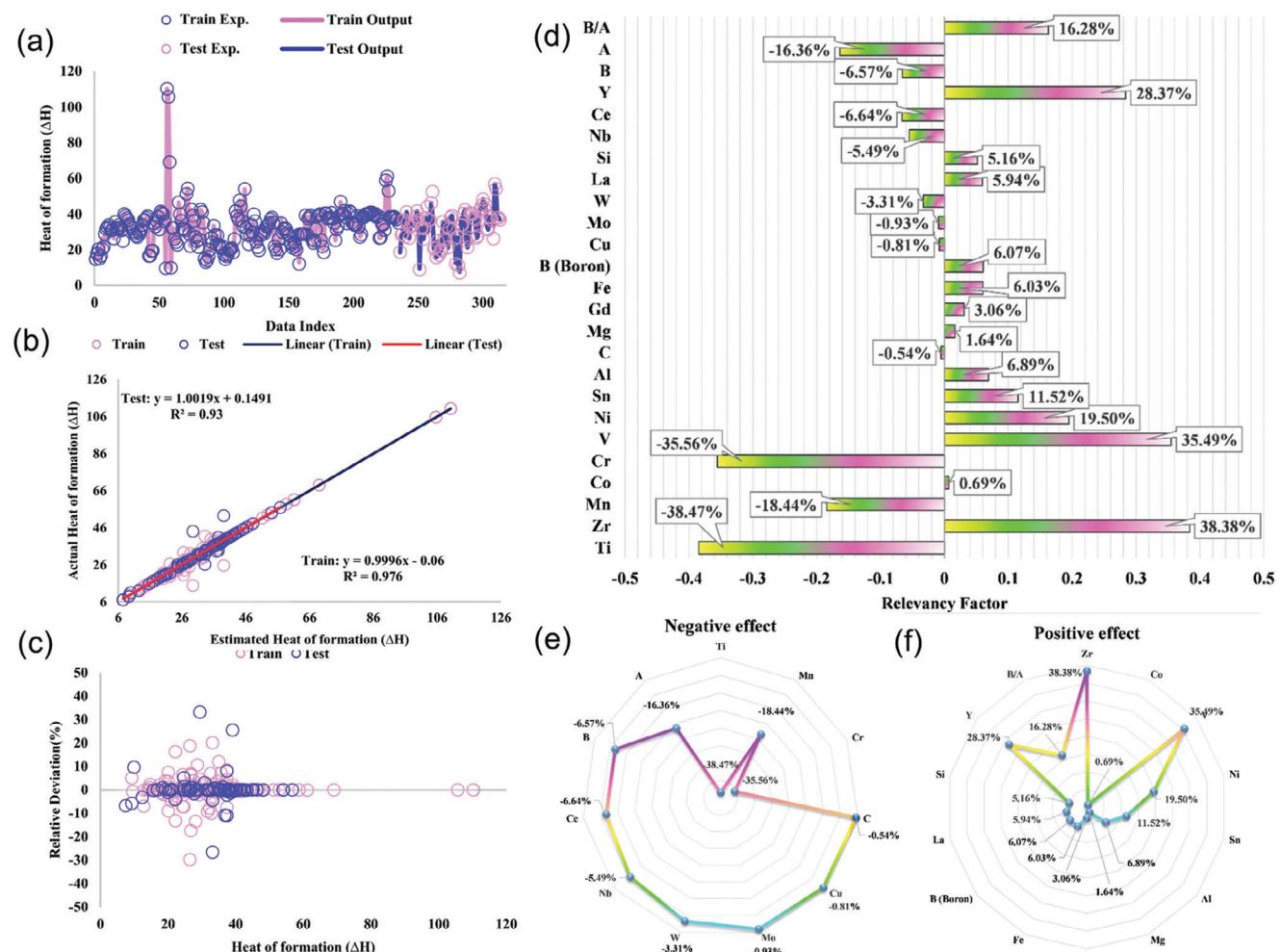
In addition, the integration of hydrogen storage capacity and hydride formation enthalpy can also contribute to the design of novel materials. Utilizing the HydPARK dataset alongside Magpie descriptors, Witman et al.<sup>[27c]</sup> discovered a clear relationship between the de-/hydrogenation equilibrium pressure and volume-based descriptors, as evidenced by feature importance analysis of GBDT model ( $\text{MAE}_{\text{train}} = 0.47$ ;  $\text{MAE}_{\text{test}} = 1.52$ ). This finding emphasizes the intrinsic link among the hydride formation enthalpy, metal hydride stability, and the cell volume factor (Figure 5a-d). By conjugating this insight, they charted the course for the design of innovative high-pressure AB<sub>5</sub>-type HSM (UNi<sub>5</sub>) with low thermodynamic stability, as successfully validated with DFT results. Hattrick-Simpers et al.<sup>[27e,f]</sup> employed ML to predict hydride formation enthalpy incorporating techno-economic analysis as a constraint, where RF model has the highest training accuracy with an MAE of  $8.56 \text{ kJ mol}^{-1} \text{ H}_2$  and an average relative error of 28%. However, ML models struggle to accurately describe the entropy value due to irregular numerical fluctuations. By imposing conditional constraints related to alloy cost, composition, and potential structures, they narrowed down the initial pool of 6110 potential alloys suitable for hydrogen compressors to less than 400. Ultimately, in collaboration with expert heuristics, the Fe-Mn-Ti-X alloy system was identified as an ideal candidate for the compositional development of future hydrogen compression materials. As shown in Figure 5e, Verma et al.<sup>[11e]</sup> proposed a ML framework titled HEART, empowering accurate predictions of hydrogen storage capacity (HYST) and enthalpy (THOR) across various temperatures, with commendable validation set perfor-

mance (HYST:  $R^2 = 0.81$ ,  $\text{MAE} = 0.45 \text{ wt\% H}_2$ ; THOR:  $R^2 = 0.89$ ,  $\text{MAE} = 4.53 \text{ kJ mol}^{-1} \text{ H}_2$ ). This research underscores the importance of incorporating diverse data and pertinent features (e.g., hydrogenation temperature, atomic interaction behaviors, and elemental properties) can enhance model training accuracy. Moreover, when alloying small proportions of non-hydrogenated elements (e.g., Cr, Fe, Cu, Ni, Co, and Sn) with hydrogenated elements (e.g., Mg, Al, V, and Ti), it results in a diminishing of enthalpy absolute value, albeit with minimal impact on hydrogen storage capacity. Through HEART framework, the successful screening of 6480 potential compositions from a vast pool of 6.4 million multicomponent alloys was achieved, meeting the rigorous static hydrogen storage criteria established by the DOE, namely, exceeding 2.5 wt%  $\text{H}_2$  capacity at room temperature and a  $\Delta H$  below  $60 \text{ kJ mol}^{-1} \text{ H}_2$ .

As previously stated, the HydPARK dataset serves as a valuable resource for identifying universal properties of solid-state HSMs and facilitating large-scale preliminary material screening. However, given the inherent variations in de-/hydrogenation characteristics among various solid-state HSMs, the extracted laws may not precisely capture the subtle differences of specific material categories. Consequently, it is imperative to undertake ML research tailored to specific HSM systems to achieve a more accurate understanding.

### 3.2. Ti-Based Hydrogen Storage Materials

Titanium-based HSMs are prime candidates for engineering applications, renowned for their satisfactory gravimetric hydrogen storage capacity (exceeding 1.80 wt%  $\text{H}_2$ ), broad



**Figure 6.** a–c) Prediction behaviors of GPR model with exponential kernel; d–f) Sensitivity analysis of different alloying element to hydride formation enthalpy.<sup>[120]</sup> a–f) Reproduced with permission.<sup>[120]</sup> Copyright 2022, Springer Nature.

modulation space for de-/hydrogenation performance, remarkable cycling durability, and cost-effectiveness, such as AB<sub>2</sub>-type C14/C15 Laves HSMs and AB-type TiFe-based HSMs.<sup>[119]</sup> At present, as exhibited in Table 3, ML studies on Ti-based HSMs focuses on AB<sub>2</sub>-type C14 Laves HSMs, specifically including phase structure, hydride formation enthalpy, room temperature de-/hydrogenation performance, hydrogen storage capacity, PCT curves, etc.

Regulation of crystal structure and thermodynamic de-/hydrogenation performance is the emphasis in the study of Ti-based Laves-type HSMs. Taking Zr-Ti-based Laves-type HSMs which were reported between 1998 and 2019 as the research subject, Suwarno et al.<sup>[32]</sup> investigated the effects of alloying elements on the hydrogen storage performance with ML approaches. In specific, RF model was proved to be relatively adept at predicting target values such as hydride formation enthalpy ( $R^2 = 0.647$ ), phase abundance ( $R^2 = 0.832$ ), and hydrogen storage capacity ( $R^2 = 0.688$ ). Insights gained from the visualized decision-making process reveal that Ni contributes the most significantly to the variation in hydride formation enthalpy but reduces hydrogen storage capacity. Meanwhile, Cr and Mn elements play crucial

roles in promoting the formation of C14-type Laves phase and enhancing hydrogen storage capacity, respectively. As depicted in Figure 6a–c, Gheytanzadeh et al.<sup>[120]</sup> investigated the effects of alloying elements on the hydride formation enthalpy with gaussian process regression (GPR) and exponential kernel ( $R^2 = 0.969$ ; MSE = 3.909 kJ mol<sup>-1</sup> H<sub>2</sub>). In accordance with the sensitivity analysis, Ti, Zr, Cr, and V elements have the greatest influence on the hydride formation enthalpy, thereby effectively modulating the thermodynamic stability of hydrides (Figure 6d–f). To describe the hydrogen storage capacity, Maghsoudy et al.<sup>[121]</sup> achieved high-precision predictions ( $R^2 = 0.980$ ; MSE = 0.0020; RMSE = 0.045) using the GA-PSO (Particle Swarm Optimization, an optimization algorithm that simulates bird swarm foraging and seeks optimal solutions via particle information sharing & collaboration)-Least Square SVM algorithm. According to the sensitivity analysis conducted, the contents of Sn, Co, and Ni elements, with correlation coefficients of 44.65%, 34.67%, and 34.03% respectively, have the most significant impact on the hydrogen storage capacity.

With a view to achieving optimization of the room temperature hydrogen storage performance for Ti-based HSMs,

Dangwal et al.<sup>[33]</sup> optimized a high-performance GPR model to predict the target value of hydride formation enthalpy with  $R^2$  reaching 0.98 based on exponential kernel and self-validation method with outlier removal. The hydride formation enthalpy of  $Ti_xZr_{2-x}CrMnFeNi$  high entropy alloy (HEA) ( $x = 0.5, 1.0, 1.5$ ) predicted by the GPR model is basically consistent with the experimental results. Among them, the predicted hydride formation enthalpy of  $TiZrCrMnFeNi$  HEA is  $-27.3 \text{ kJ mol}^{-1} H_2$ , with an error of only 5% compared to the experimental value of  $-28.8 \text{ kJ mol}^{-1} H_2$ . In addition, Zhou et al.<sup>[11d]</sup> employed implicit/explicit feature-based ML for the first time in the study of structure and room temperature hydrogen storage performance of C14-Laves single-phase alloys, in which SVM model can well describe the target values of cell volume and de-/hydrogenation equilibrium pressures ( $R^2 > 0.96$ ), while GBDT model has better performance to the target values of hydride decomposition enthalpy and hydrogen storage capacity ( $R^2 > 0.80$ ), as displayed in Figure 7a–e. In combination with the feature importance analysis and DFT calculation, the crucial factors affecting the hydrogen storage capacity (MeanIonicChar and Fe content) and the positive correlation between the atomic occupation disorder and the slope of the PCT plateau were revealed (Figure 7f–h). Subsequently, based on the active composition-property scanning of ML model and following targeted parameter screening (Figure 7i–k),  $Ti_{0.9}Zr_{0.12}Mn_{1.2}Cr_{0.55}(VFe)_{0.25}$  alloy with cost-effective and comprehensive performance has been successfully customized for the proton exchange membrane fuel cell hydrogen source system.

PCT curves serve as a straightforward source for obtaining the thermodynamic de-/hydrogenation performance of HSMs and evaluating their overall performance. To acquire high-precision PCT curve prediction models, Kim et al.<sup>[122]</sup> generated substantial additional data points by a combination of PCT data point fitting and Van't Hoff equation projection. Among the optimized models (RF, KNN, and DNN), DNN model has the best performance with an average correlation coefficient of 0.93070, which can successfully predict the de-/hydrogenation PCT curves for different alloy compositions at arbitrary temperatures. Thus, high-accuracy ML model can offer valuable insights for anticipating and fine-tuning the de-/hydrogenation behaviors of solid-state HSMs.

Small alterations in elemental composition are sufficient to alter the atomic arrangement and hydrogen storage performance of solid-state HSMs. Presently, Ti-based HSMs predominantly consist of conventional elements such as Ti, Zr, Mn, Cr, V, and Fe, with limited experimental and ML research focusing on incorporating novel elements to fine-tune the alloy's phase structure and overall hydrogen storage performance. Besides, ML investigations on C15 Laves-type and AB-type Ti-based HSMs have yet not be conducted, suggesting there are significant untapped opportunities and potential for developing innovative high-performance Ti-based HSMs.

### 3.3. Rare Earth-Based Hydrogen Storage Materials

In view of the abundance of rare earth resources, superior reversible de-/hydrogenation performance, high volumetric hydrogen storage capacity ( $>120 \text{ kg m}^{-3}$ ), excellent anti-poisoning

properties and cycling durability, rare earth (RE)-based HSMs are prime candidates for solid-state hydrogen storage applications, and specifically include  $AB_5$ -type HSMs and superlattice HSMs.<sup>[123]</sup> Currently, as exhibited in Table 3, ML research towards RE-based HSMs includes aspects of chemical/electrochemical hydrogen storage performance.

As for chemical hydrogen storage applications, Zhao et al.<sup>[124]</sup> explored the factors affecting the hydride formation enthalpy and hydrogen storage capacity of  $LaNi_{5-x}M_x$  ( $M = Si, Al, Mn, Fe, Co, Ni, Ga, and Ge$ ) system through atomic parameters and stepwise regression. For a given system, the lower the atomic size and electron concentration, the more pronounced the difference in electronegativity, the higher the hydride formation enthalpy, indicating the greater the thermodynamic stability of metal hydride at a given temperature. Besides, the reduction in the values of parameters such as atomic size difference, electronegativity difference, electron concentration, and hydrogenation temperature of the alloys is favorable for interstitial hydrogen atom occupancy and larger hydrogen storage.

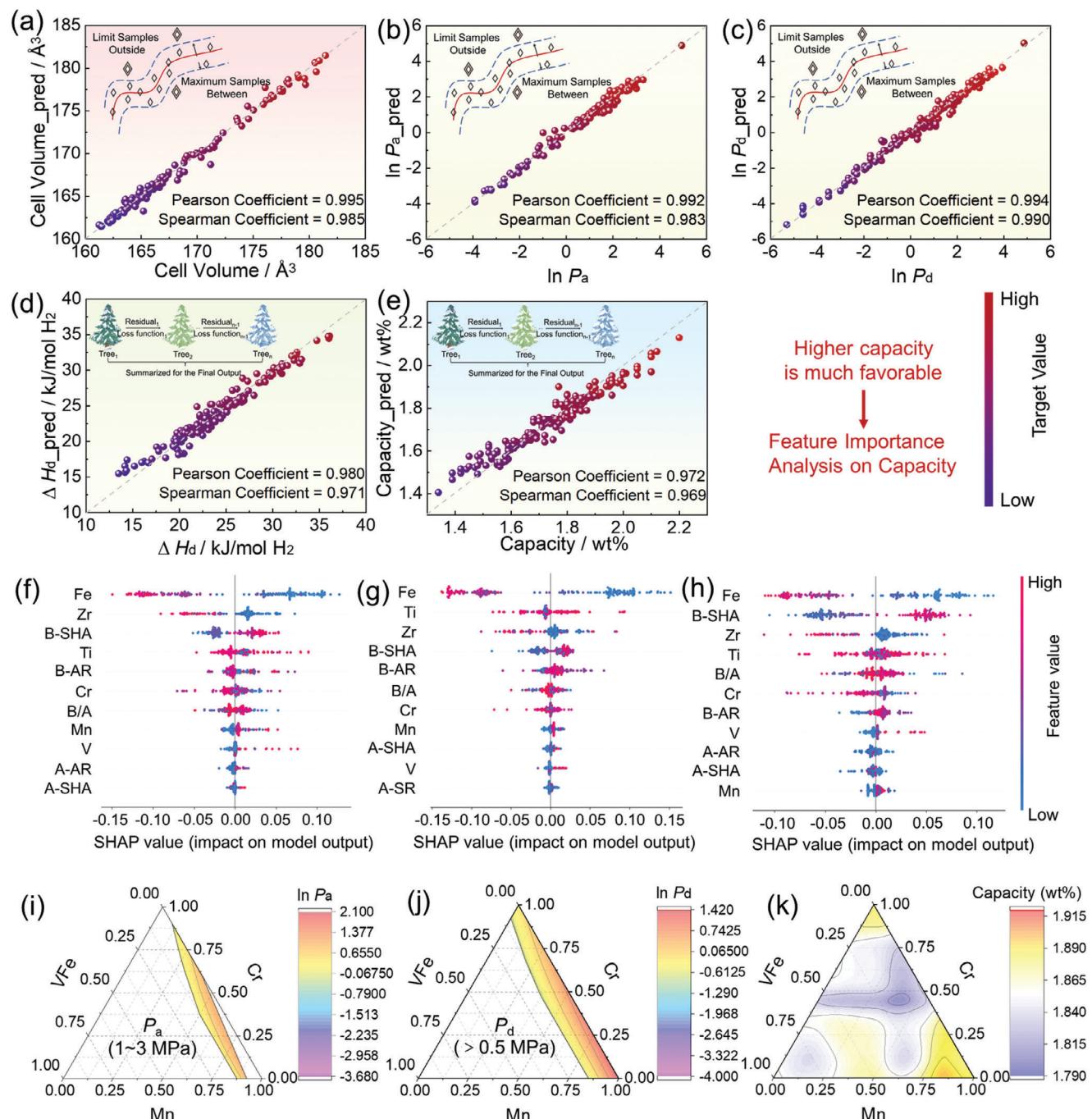
With respect to electrochemical hydrogen storage applications, Zhao et al.<sup>[125]</sup> explored the influence of various substitutional elements on the electrochemical hydrogen storage performance of  $(La, Ce, Pr, Nd)_2MgNi_9$  alloy based on the SVR model and PSO method. With the addition of Ce, Pr, and Nd elements, the fast discharge capability of the alloys can be significantly enhanced although it is detrimental to the maximum discharge. The results indicate that the favored optimized alloy with the maximum discharge capacity is  $(La_{1.633}Ce_{0.059}Pr_{0.249}Nd_{0.59})_2MgNi_9$ , whereas  $(La_{1.44}Ce_{0.06}Pr_{0.25}Nd_{0.25})_2MgNi_9$  possesses the best fast discharge performance.

As aforementioned, current ML study on rare earth-based HSMs predominantly focuses on the elemental/surface modification effects, whereas the systematic high-throughput modulation of the compositional properties and the development of novel materials have not yet been reported, necessitating the continuous follow up of the subsequent studies. In addition, it is also worthwhile to conduct ML studies on the subunits synchronous de-/hydrogenation modulation and cycling durability enhancement of RE-based superlattice HSMs.

### 3.4. Solid Solution Hydrogen Storage Materials

Based on the non-close-packed stacking characteristics, solid solution HSMs could accommodate three times as many interstitial sites per unit cell as hexagonal close packed (HCP) structures, and thus attain higher hydrogen storage capacity ( $H/M \geq 2$ ).<sup>[126]</sup> So far, ML research for solid solution HSMs mainly focuses on hydrogen storage capacity, hydride formation/decomposition enthalpy, phase diagram, new materials mining, etc.

Alloying elements determine the amount of hydrogen stored in solid solution HSM by influencing the interstitial size and hydrogen storage environment. Lu et al.<sup>[127]</sup> employed ensemble model to precisely predict the saturated hydrogen storage capacity of V-Ti-Cr-Fe-based HSMs, achieving a MSE value of 0.187. According to feature importance ranking, it's evident that valence electron concentration, lattice constant, and  $Z/r^3$  (valence electron/atomic radius) exert significant effects on the capacity. Additionally, the optimal composition ranges of three high-capacity



**Figure 7.** a–e) Prediction performance of optimal ML model (SVM or GBDT) for key structure and hydrogen storage performance; f–h) SHAP analyses for hydrogen storage capacity; i–k) Predicted results with averaging ML models for the novel  $Ti_{0.9}Zr_{0.12}Mn_xCr_y(VFe)_{2-x-y}$  ( $0 \leq x, y \leq 2$ ,  $0 \leq x+y \leq 2$ ) alloy system.<sup>[11d]</sup> a–k) Reproduced with permission.<sup>[11d]</sup> Copyright 2023, Elsevier Ltd.

alloys are further determined with genetic algorithm (GA, iteratively optimizes solutions by modeling natural selection and genetics), and experimentally verified that the relative errors of their predictions are  $\approx 1\%$ , demonstrating the generalization ability of the model. Likewise, through SVR analysis, Zhao et al.<sup>[125]</sup> discovered that in  $Ti_{0.32}Cr_{0.43-x-y}V_{0.25}Fe_xMn_y$  solid solution HSMs system, the influence of Mn on the effective hydrogen storage capacity is contingent upon the Fe content. Specifically, an abundance of Fe can deteriorate the high-capacity Laves phase structure, thereby necessitating a low Fe content for Mn additions to effectively enhance the hydrogen storage capacity.

It is a common approach to modulate the thermodynamic stability of solid solution hydrides through modulating the hydride formation/decomposition enthalpy. Lu et al.<sup>[38b]</sup> predicted the hydride decomposition enthalpy of V-Ti-Cr-Fe alloys using the XG-Boost model, achieving the relative error below 15%. Afterward,

the relative discrepancies between the actual hydride decomposition enthalpy and the predicted value of the alloy compositions that are optimized by GA are within a tight margin of 4.5%. A comprehensive analysis combining feature importance assessment and DFT calculations indicates a clear trend: the thermodynamic stability and decomposition enthalpy of the hydrides increase significantly with increasing cell volume and valence electron concentration, and decreasing bulk modulus. Similarly, Agafonov et al.<sup>[128]</sup> adopted GBDT model ( $MAE = 4.2 \text{ kJ mol}^{-1} \text{ H}_2$ ;  $R^2 = 0.90$ ) to describe the hydride decomposition enthalpy of  $(\text{TiVNb})_{75}\text{Cr}_{25-x}\text{Fe}_x$  ( $0 < x < 20$ ) alloys. Attributed to the lower hydrogen affinity of the Fe element, their conclusions indicate that as the Fe content rises, there is a corresponding decrease in both the hydride decomposition enthalpy and metal hydride stability. Moreover, Witman et al.<sup>[28]</sup> screened high entropy solid solution HSMs with potential thermodynamic instabilities, such as  $\text{AlTiVNbTa}$  and  $\text{AlTiVCr}$ , based on the GBDT model ( $\ln(P_{eq}^\circ/P_0)$ - $MAE: 1.4$ ;  $\Delta H$ - $MAE: 5.5 \text{ kJ mol}^{-1} \text{ H}_2$ ). Compared to the benchmark  $\text{TiVZrNbHf}$  HSM, these HSMs have higher  $\ln(P_{eq}^\circ/P_0)$  values and contain lighter and less-costly Al element. Experimentally prepared single-phase HSMs give measured values in general agreement with predictions, providing a solution for the development of lighter and thermodynamically more unstable metal hydrides.

The phase diagram of metal-hydrogen systems visually encapsulates the intricate de-/hydrogenation behavior and phase transformation boundaries exhibited by solid-state HSMs. With the dataset derived from DFT calculations, Witman et al.<sup>[129]</sup> devised an efficient approach to model the phase diagrams of alloys and their hydrides, taking advantage of the power of GNNs even with limited training data. In the pursuit of optimizing thermodynamic stability, a pivotal insight emerged: the elusive synthesis of the pure-phase  $\text{AlLiMgSnZn}$  solid solution in experiments can be traced back to the exorbitantly free energy of the target phase, exceeding the convex hull threshold, thereby inducing its instability and decomposition. Accordingly, strategic adjustments such as reduction or elimination of Sn element can significantly enhance the thermodynamic stability of the alloy system (Figure 8a). Exploiting this innovative model, energy scanning and determination of Pd-substituted alloys were performed. As depicted in Figure 8b–g, this is the first time that PCT curves can be computed with unprecedented precision, enabling a quantitative assessment of how alloy substitutions modulate the PCT platform pressure and saturation capacity. This breakthrough not only illuminates the intricate phase boundaries that govern the metal-to-hydride phase transformation, but also highlights the potential of computational methods in advancing hydrogen storage technology.

Limited by the high thermodynamic stability of metal hydrides and the unique de-/hydrogenation characteristics of dual plateaus, it's a challenging mission for solid solution HSMs to realize reversible de-/hydrogenation processes and high reversible capacity utilization under moderate conditions. Accordingly, Halpren et al.<sup>[40b]</sup> demonstrated that the hydrogenation thermodynamics of the first and second stages of high-entropy solid solution HSMs are primarily governed by the bulk elastic modulus and *d*-band center, respectively, through interpretable ML models and DFT calculations, as depicted in Figure 9a–f. In this work, the Multi-Objective Bayesian Optimization (MOBO) technique

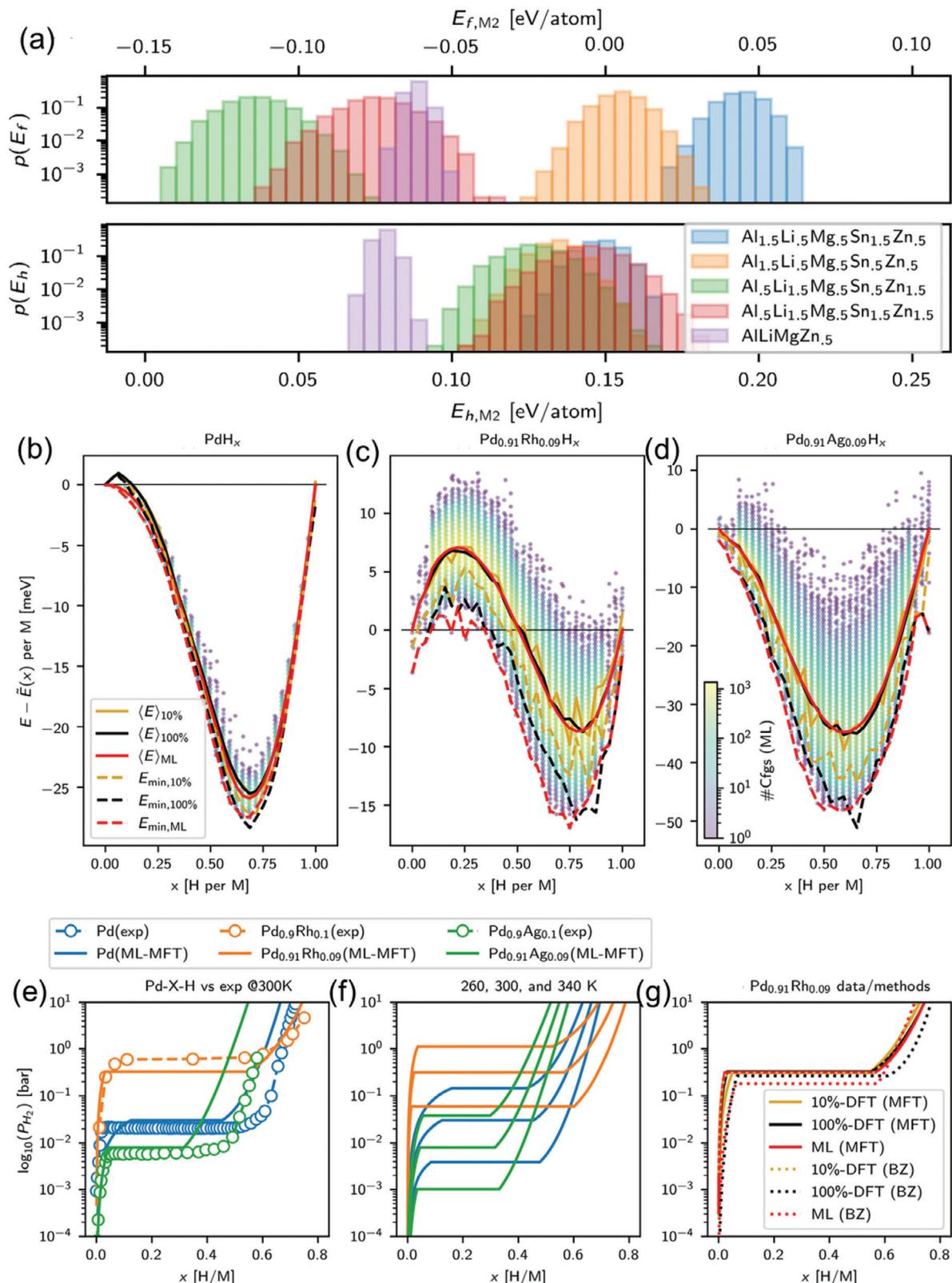
and extended DFT dataset were employed to explore and balance the hydrogen binding energies for the first ( $\Delta E_{mono}$ ) and second ( $\Delta E_{di}$ ) stages of hydrogenation reactions within the intricate compositional landscape of BCC high-entropy solid-state HSMs. Among the eight novel HSM candidates identified through this optimization strategy, the optimized  $\text{VNbCrMoMn}$  alloy exhibits the most favorable hydrogenation behaviors, with  $\Delta E_{mono}$  and  $\Delta E_{di}$  values of  $-0.129 \text{ eV}$  and  $-0.131 \text{ eV}$ , respectively, which are close to the desirable  $-0.12 \text{ eV}$  (Figure 9g). As a result, it is capable of a theoretical hydrogen storage capacity of  $2.83 \text{ wt\% H}_2$  at room temperature and ambient hydrogenation pressure, significantly surpassing that of commercial  $\text{LaNi}_5\text{H}_6$  ( $1.38 \text{ wt\% H}_2$ ) and  $\text{TiFeH}_2$  ( $1.91 \text{ wt\% H}_2$ ).

Compared to Ti-based and RE-based HSMs, solid solution HSMs exhibit superior saturated hydrogen storage capacity. However, they struggle to attain high reversible de-/hydrogenation efficiency at moderate conditions due to their dual-platform characteristics and high stability. Presently, the regulation of their room temperature hydrogen storage performance relies solely on theoretical calculations and lacks experimental evidence. Hence, a synergistic approach combining both theoretical and experimental endeavors is indispensable for the modulation of de-/hydrogenation pathways.

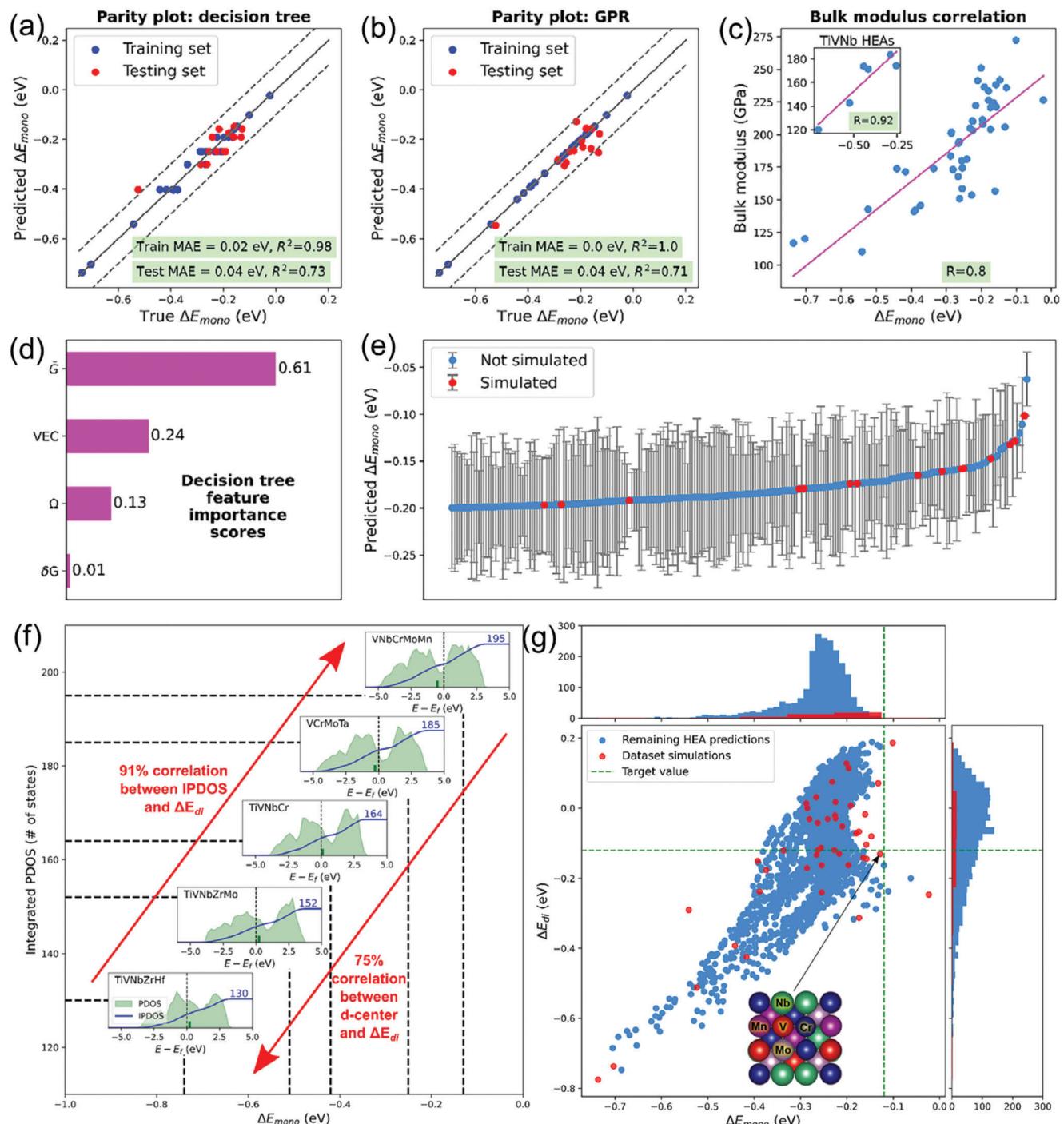
### 3.5. Mg-Based Hydrogen Storage Materials

Mg-based HSMs are promising candidates for solid-state HSMs by virtue of their high hydrogen storage capacity, broad resource base, and low cost.  $\text{MgH}_2$  reaches the hydrogen storage capacity of  $7.60 \text{ wt\% H}_2$ , but its high de-/hydrogenation temperatures and sluggish de-/hydrogenation kinetic rates severely limit its cycling efficiency.<sup>[130]</sup> Currently, modification techniques including multi-alloying, catalyst doping, and nano-confinement have reached a high level of maturity, enabling proficient modulation of microstructures, enhancement of the surface activity, and optimization of the reaction pathways.<sup>[131]</sup> As illustrated in Table 3, ML research on Mg-based HSMs in the literature encompasses the influence of elemental composition on de-/hydrogenation performance, cycling capacity, formation energy, interatomic interactions, new material design, etc.

Alloying modification, as an effective means of bulk phase modulation, can remarkably boost the de-/hydrogenation performance of Mg-based HSMs. Malinova et al.<sup>[132]</sup> innovatively combined the Levenberg-Marquardt training algorithm with Bayesian regularization to construct the ANN model. This two-layer feedforward hierarchical model achieved remarkable accuracy in predicting hydrogen storage capacity, dehydrogenation temperature, and plateau pressure of  $\text{MgH}_2$  with varying alloying elements ( $R > 0.95$ ). The findings reveal that barring the Al element, all 23 remaining alloying elements contribute to a decline in the hydrogen storage capacity of  $\text{MgH}_2$ . The inclusion of Ti, Ce, Y, Si, or Mm element reduces the dehydrogenation temperature and increases the dehydrogenation pressure. In contrast, the introduction of Li or Al element raises the dehydrogenation temperature while lowering the plateau pressure. Interestingly, substituting with La, Cu, Zn, Fe, or Sn can simultaneously elevate both the dehydrogenation temperature and plateau pressure. Notably, Ni is the only element that reduces both



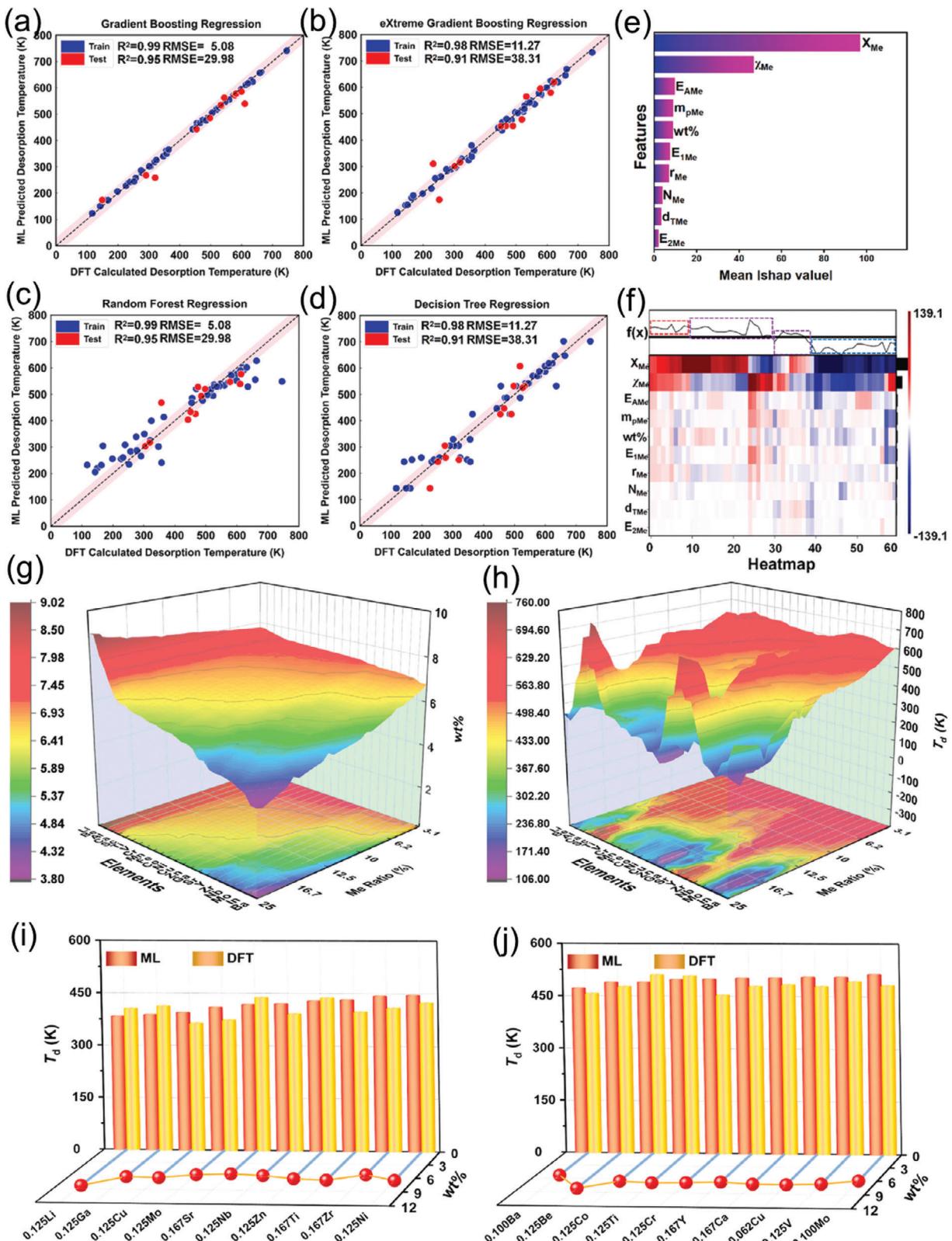
**Figure 8.** a–c) Configuration energy as well as the average energy ( $\langle E \rangle(x)$ ) and minimum energy ( $E_{\min}(x)$ ) with 10%, 100% training data, and ML model; d) PCT curve of Pd-X-H at 300 K; e) PCT curves with increasing temperature; f) PCT curve comparisons with different prediction approaches.<sup>[129]</sup> a–g) Reproduced with permission.<sup>[129]</sup> Copyright 2024, American Chemical Society.



**Figure 9.** Parity plots of  $\Delta E_{mono}$  with a) DT and b) GPR; c) The relationship between bulk modulus and  $\Delta E_{mono}$ ; d) Feature importance scores by decision tree; e) The uncertainties of  $\Delta E_{mono}$  predictions; f) The relationship between d-center and  $\Delta E_{di}$ ; g) The distribution of predicted and simulated hydrogen binding energy for each absorption phase by using the corresponding GPR model.<sup>[40b]</sup> a–g) Reproduced with permission.<sup>[40b]</sup> Copyright 2024, Elsevier Ltd.

dehydrogenation temperature and plateau pressure. Likewise, Jiang et al.<sup>[133]</sup> explored the effect of different alloying elements on the dehydrogenation temperature of  $MgH_2$  with GBDT model ( $R^2 = 0.95$ ; RMSE = 29.98 K), shown in Figure 10a–d. Through SHapley Additive explanation (SHAP) analysis, it is clearly

discernible that the doping ratio and electronegativity of alloying elements have the greatest impact on the dehydrogenation temperature of the modified  $MgH_2$ . Furthermore, as the alloying proportion increases, the extent of increase in the dehydrogenation temperature diminishes (Figure 10e,f). Also, among the



**Figure 10.** a–d) Parity plots of desorption temperature with different tree models; e) SHAP value evaluation of the features; f) SHAP matrix heatmap; Distribution of g) hydrogen storage capacity and h) dehydrogenation temperature for substituted  $\text{MgH}_2$ ; DFT calculations and ML predictions of dehydrogenation temperature and hydrogen storage for substituted  $\text{MgH}_2$  at i) 400 K and j) 500 K.<sup>[133]</sup> a–j) Reproduced with permission.<sup>[133]</sup> Copyright 2024, Elsevier Ltd.

evaluated elements, the addition of Li and Be is accompanied by an increase in the hydrogen storage capacity while lowering the dehydrogenation temperature, and the obtained  $Mg_{0.875}Li_{0.125}H_2$  and  $Mg_{0.875}Be_{0.125}H_2$  are the ones with the highest hydrogen storage capacity at 400 K and 500 K, respectively (Figure 10g–j).

Mg-based HSMs are a promising type of cathode material for Ni-MH batteries due to their high discharge capacities, abundant sources, and low cost. However, their major drawback lies in their limited cycle life. To address this shortcoming, Tian et al.<sup>[134]</sup> proposed a cycle lifespan prediction model supported by BPNN, which can precisely estimate the normalized discharge capacity for cycles beyond the fifth of Mg-Ti-M-Ni-based alloys ( $M = Ti, Al, Cr, etc.$ ) and Mg-Ti-Pd-Ni-based alloys, respectively ( $R^2 > 0.99$ ). Additionally, cycle lifetime of  $Mg_{0.9}Al_{0.08}Ce_{0.02}Ni$  alloy and  $Nd_5Mg_{41}-Ni$  composite can both be successfully predicted by the model with  $R^2$  values exceeding 0.996, justifying its robustness and precision in cycle lifespan estimations of Mg-based HSMs regarding electrochemical performance.

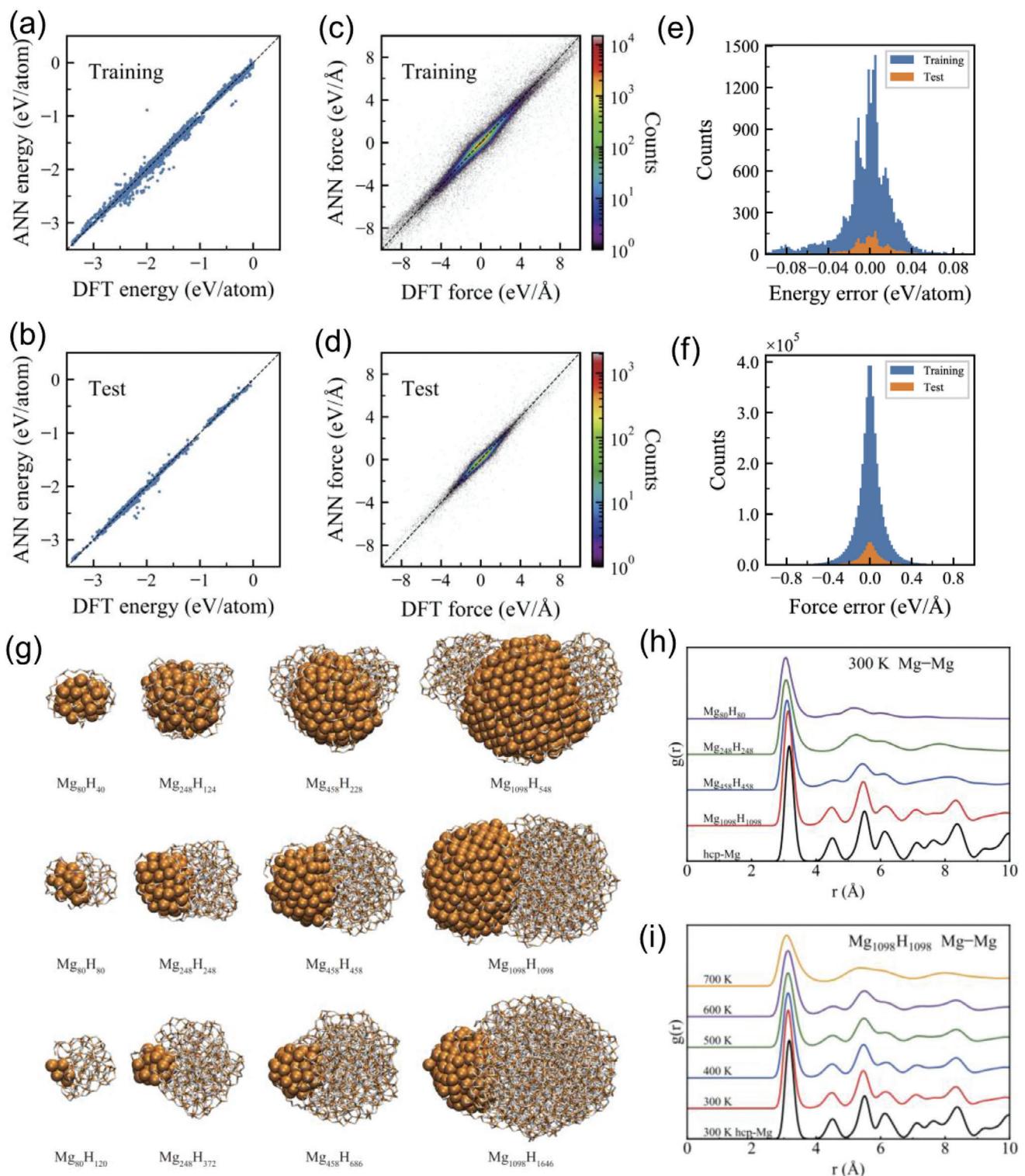
Thermodynamic stability is the primary prerequisite for screening solid-state HSMs. He et al.<sup>[34]</sup> utilized two methods, kernel ridge regression (KRR) and deep potential molecular dynamics (DeePMD, a MD simulation framework that utilizes DFT data and NN to generate multi-body potentials and atomic force fields, all without the need for human intervention), to develop ML models aimed at predicting the formation energy of Mg-based HSMs. The results reveal that DeePMD model excelled in terms of prediction accuracy (RMSE = 0.43 meV per atom), while KRR model demonstrates superior time efficiency with an RMSE of 6.80 meV/atom. Both models can well estimate the formation energy of binary Mg-based HSMs within milliseconds, maintaining an RMSE range of 0–10 meV/atom when compared to that from DFT calculations. Thereby, these models offer a precise and efficient approach for predicting and designing the formation energy of Mg-based HSMs.

Since DFT is difficult to achieve the calculation of structure and kinetic properties of large-sized hydride nanoclusters, Wang et al.<sup>[135]</sup> developed a neural network interatomic potential applicable to the Mg-H system. Specifically, they adopted first-principles calculation result as the data source, and the comprehensive discussion covers several key aspects: the construction/optimization of the NN model, the strategic design of the sampling process, the rigorous error evaluation, and a thorough examination of applicability and performance.<sup>[135]</sup> As displayed in Figure 11a–f, the energy and force predicted by ANN are consistent with DFT calculations at any Mg/H ratio (energy-MAE: 18.23 meV per atom; force-MAE = 95.48 meV Å<sup>-1</sup>). Besides, in the case of hydrogen-saturated clusters ( $n_{Mg}: n_H = 1:2$ ), the atoms are essentially disordered at 300 K. Instead, hydrogen-unsaturated clusters ( $n_{Mg}: n_H < 1:2$ ) are prone to generate Mg/MgH<sub>x</sub> dual-phases, and at 600 K, the Mg phase therein even forms ordered nanocrystalline structures (Figure 11g–i). It is evident from MD simulations that hydrogen diffusion displays a temperature dependency between 400 K–700 K, but is independent of the hydrogen content of the clusters. Based on this, Wang et al.<sup>[136]</sup> combined force-based DFT with ANN and found that Ni atoms can weaken the coulomb repulsion of surrounding atoms and reduce the bonding strength with hydrogen atoms. What's more, the five-coordinated Ni site can act as an intermediate species during the dehydrogenation process

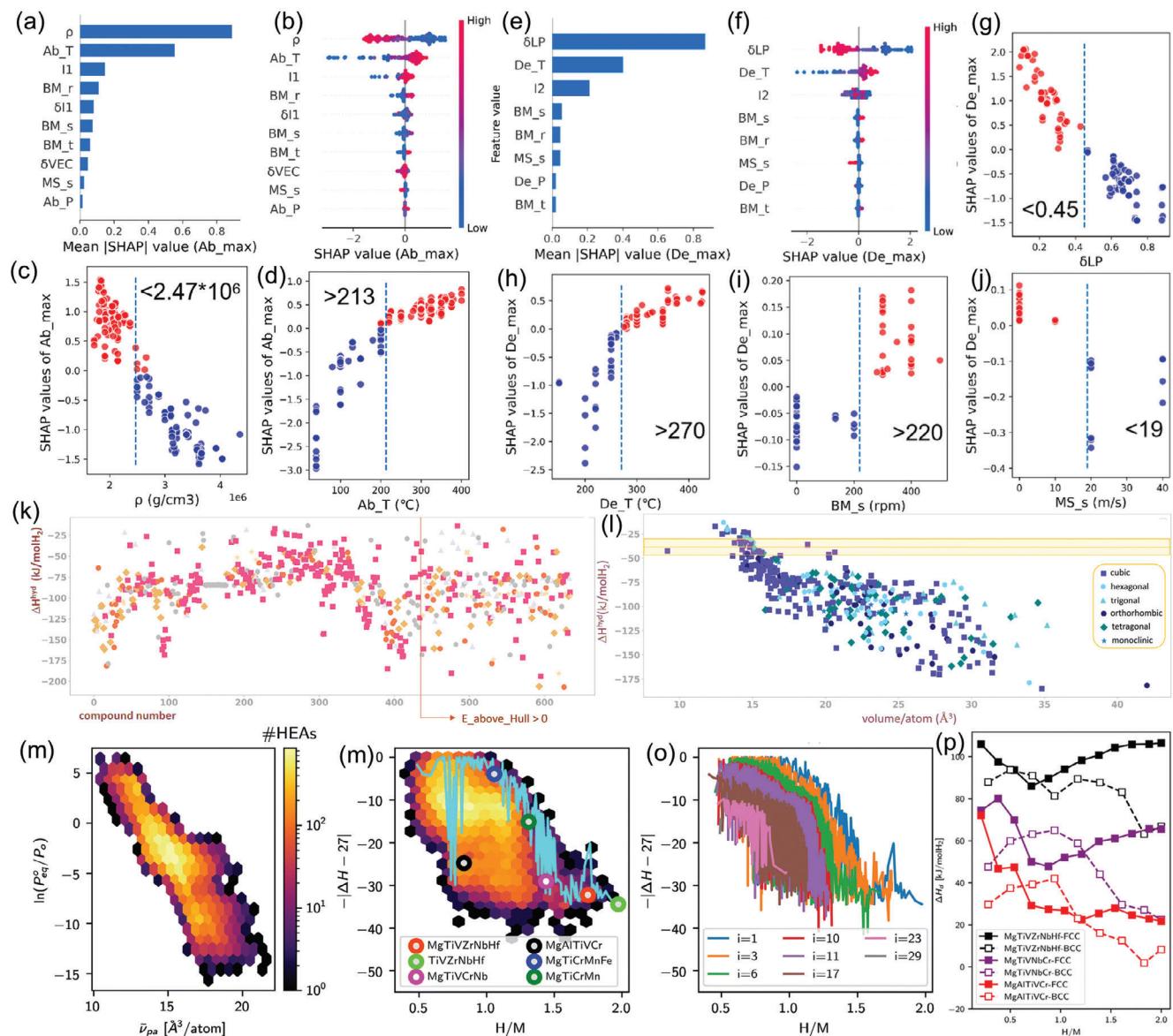
and provides a fast channel for hydrogen diffusion, and thus Ni doping can significantly enhance the dehydrogenation properties of Mg-based metal hydride clusters. To accurately predict the dehydrogenation performance of large-sized  $Mg_nH_{2n}$  nanoparticles, Pedrielli et al.<sup>[137]</sup> initiated an investigation into the thermodynamic properties of  $Mg_nH_{2n}$  NPs ( $n < 10$ ) with DFT calculations, particularly evaluating the effects of anharmonicity on enthalpy, entropy, and thermal expansion through the Stochastic Self-Consistent Harmonic Approximation (SSCHA). Their findings uncovered a nearly linear correlation between interatomic bond lengths and temperature, with minimal variations of just a few percent above 300 K, alongside a reduction in Mg-H bond length. To amplify the size effect in nanoparticles, they devised a ML algorithm, SchNet-SSCHA (S[chnet]SCHA), employing the SchNet NN package (energy-MAE: 0.02 eV; force-MAE = 0.02 eV Å<sup>-1</sup>). This innovative model can predict both the forces and total energy (potential energy surfaces) of nanoparticles, significantly advancing the stochastic self-consistent harmonic approximation (SSCHA) method. This approach is particularly advantageous for the study of  $Mg_nH_{2n}$  nanoparticles ( $n$  up to 43), as it allows for the exploration of their thermodynamic properties and free-energy landscape at a fraction of the computational cost associated with traditional ab initio methods.

In addition, to investigate the feasibility of two-dimensional Mg-based hydrides ( $Mg_xH_y$ ) for application, Lee et al.<sup>[138]</sup> applied a crystal inverse design technique based on generative adversarial networks (GANs, structure validation ratio: 96.8%; structure generation ratio: 87.3%). They constructed a 2D binary phase diagram to determine crucial crystal configurations. Through DFT calculations, they delved into the electronic and dynamic characteristics of  $Mg_xH_y$  sheets within low-energy periodic phases, and ultimately discovered a novel 2D MgH<sub>2</sub> phase belonging to the  $P\bar{4}m2$  space group. Further exploration reveals that lithium-modified MgH<sub>2</sub> offers remarkable hydrogen storage capacity up to 6 wt% H<sub>2</sub> coupled with stable absorption energy, highlighting its potential as an efficient hydrogen storage carrier. Notably, the incorporation of lithium can significantly boost the absorption capacity of hydrogen molecules on the surface of MgH<sub>2</sub> surface. Moreover, the thermal stability of this lithium-decorated MgH<sub>2</sub> at ambient temperature was confirmed by AIMD simulations, indicating its potential feasibility for future applications.

New material design is an attractive research branch of ML for Mg-based HSMs, including design guideline extraction and high-throughput screening. Dong et al.<sup>[139]</sup> achieved precise predictions of the maximum absorption capacity (Ab\_max) and desorption capacity (De\_max) of Mg-based HSMs using GBDT and MLP models, with testing set  $R^2$  values reaching 0.947 and 0.922, respectively. As displayed in Figure 12a–j, the threshold values were identified by the positivity or negativity of SHAP values (Ab\_max: density( $\rho < 2.47 \times 10^6$  g cm<sup>-3</sup>) and hydrogenation temperature (Ab\_T > 213 °C); De\_max: lattice parameter mismatch ( $\delta LP < 0.45$ ), dehydrogenation temperature (De\_T > 270 °C), milling speed (BM\_s > 220 rpm), and melt spinning rate (MS\_s < 19 m s<sup>-1</sup>)), supplying necessary guidelines for the design of high-performance Mg-based HSMs. Employing the optimized ML model, the Ab\_max and De\_max of Mg-based binary alloys with 16 additional metallic elements and MgNi-based ternary alloys with 15 additional metallic elements were predicted. Among which, 96Mg-4Sm and 95Mg-1Ni-4Sm exhibit



**Figure 11.** a-d) Parity plots of energy and force with training set and test set; Counts of e) energy error and f) force error; g) Equilibrium geometry of hydrogen-unsaturated nanoclusters after 0.1 ns MD simulation at 300 K; Mg-Mg partial RDFs of atomic Mg portions in h)  $Mg_n H_n$  clusters with different sizes at 300 K and i)  $Mg_{1098}H_{1098}$  clusters at different temperatures.<sup>[135]</sup> a-i) Reproduced with permission.<sup>[135]</sup> Copyright 2020, American Physical Society.



**Figure 12.** MAE of SHAP values for selected features in a) Ab\_max dataset and e) De\_max dataset; SHAP interpretation plots for b) Ab\_max dataset and f) De\_max dataset; c,d) SHAP values for the first two features in (a,b); g-j) SHAP values of the first four features in (e,f).<sup>[139]</sup> Figure 12a–j were reproduced with permission.<sup>[139]</sup> Copyright 2023, Elsevier Ltd. k) Predicted hydride formation enthalpy for 636 Mg-containing intermetallic compounds; l) Predicted hydride formation enthalpy as a function of atomic volume.<sup>[35]</sup> k,l) Reproduced with permission.<sup>[35]</sup> Copyright 2023, Elsevier Ltd. m) Predicted equilibrium pressure screened from 17920 peculiar equimolar HEAs against  $\bar{\nu}_{pa}$ ; n) Projections of 4-dimensional Pareto front onto a 2-dimensional plot; o) Evolution of the  $i^{th}$  Pareto front after removing all materials from the  $i-1^{th}$  Pareto front materials; p) DFT-computed hydride decomposition enthalpy for MgTiVZrNbHf, MgTiVNbCr, and MgAlTiVCr.<sup>[30b]</sup> m-p) Reproduced with permission.<sup>[30b]</sup> Copyright 2023, Royal Society of Chemistry.

superior performance, with Ab\_max and De\_max of 6.31 wt% H<sub>2</sub> and 5.69 wt% H<sub>2</sub>, respectively, and 6.64 wt% H<sub>2</sub> and 5.63 wt% H<sub>2</sub>, respectively, fulfilling the demands for efficient reversible hydrogen storage below 300 °C. Based on the MatErials graph network (MEGNet), Batalović et al.<sup>[140]</sup> designed a prediction model for hydride formation enthalpy with transfer learning, grounded on the crystal structure of the initial metal/intermetallic compound (MAE = 5.5 kJ mol<sup>-1</sup> H<sub>2</sub>). The method boasts fast prediction speed and ease of operation, as it only requires the crystal structure of the starting alloy to determine the hydride formation enthalpy. Furthermore, this model is effective in distinguishing

polymorphs, namely HSMs with same chemical composition, which is beneficial for practical material performance screening. Accordingly, it's possible to screen out Mg-Ni-M ternary alloy system that may suitable for applications in moderate environmental conditions. Afterwards, Batalović et al.<sup>[35]</sup> succeeded in predicting the hydride formation enthalpy of Mg-based HSMs with previous trained GNN model,<sup>[140]</sup> and performed a high-throughput screening of 636 Mg-based materials based on their atomic compositions and optimized crystal geometries calculated by DFT. As displayed in Figure 12k,l, a total of 32 intermetallic compounds are identified as suitable near-environmental HSMs

as indicated by the hydride formation enthalpy. Furthermore, a close correlation has been observed between the volume per atom and the hydride formation enthalpy, within a range of 14.1–15.7 Å<sup>3</sup>, serving as a valuable selection criterion to pinpoint Mg-based HSMs for moderate environmental conditions.

High-entropy HSMs present an expansive composition design space, yet achieving a trade-off among diverse hydrogen storage performance (including high hydrogen storage capacity, and suitable thermodynamic stability) and enhancing comprehensive hydrogen storage performance continues to pose a formidable challenge. As illustrated in Figure 12m–o, Witman et al.<sup>[30b,141]</sup> firstly established high-accuracy GBDT models towards hydrogen storage capacity (H/M: MAE = 0.14), enthalpy (MAE = 5.4 kJ mol<sup>-1</sup> H<sub>2</sub>), entropy (MAE = 13 J mol<sup>-1</sup>/K H<sub>2</sub>), and equilibrium pressure ( $\ln P_{eq}^0/P^0$ : MAE = 0.14). Since Pareto optimization approach is an effective multi-objective optimization method to balance and optimize multiple objectives by finding Pareto optimal solutions and Pareto frontiers, it was adopted to successfully select high-performance Mg-based HSMs (MgTiVZrNbHf, MgTiVNbCr and MgAlTiVCr) from 17 920 unique equimolar HEAs for experimental synthesis and validation. ML predictions for the above materials are in good agreement with their experimental observations, and the phase transformation behaviors of the hydrogenation process from BCC structure to FCC structure were further corroborated by DFT calculations, as evidenced in Figure 12p. Such multi-objective value balancing and optimization paths provide brand new ideas for high-performance solid-state HSM screening.

As mentioned above, ML conducted on Mg-based HSMs is rather comprehensive, covering interatomic potential, formation energy, and fundamental hydrogen storage performance. Notably, the studies pertaining to ML-derived interatomic potentials offer theoretical insights and direct guidance for exploring the interatomic interaction patterns within other solid-state HSMs. Furthermore, to attain the objective of de-/hydrogenation under moderate conditions, ML applications in alloying, nano-confinement, and catalytic synergistic modifications are supposed to anticipated to expedite the development of high-performance Mg-based HSMs.

### 3.6. Complex Hydrogen Storage Materials

Complex HSMs are generally high-capacity materials (e.g., LiBH<sub>4</sub>: 18.51 wt% H<sub>2</sub>, Mg(BH<sub>4</sub>)<sub>2</sub>: 14.94 wt% H<sub>2</sub>, NaAlH<sub>4</sub>: 7.46 wt% H<sub>2</sub>, etc.) which are composed of active cations of alkali metals or alkaline earth metals and anionic ligands, and are regarded as the ideal solid-state hydrogen storage carriers.<sup>[142]</sup> In analogy with the Mg-based HSMs discussed in Section 3.5, the decomposition of stable chemical bonds confronts substantial thermodynamic energy barriers, and therefore results in high dehydrogenation temperature and sluggish dehydrogenation kinetics. At present, mature modification methods include coordination hydride complexation, catalysts, and nano-confinement.<sup>[143]</sup> However, as shown in Table 3, there are fewer studies of ML in the field of complex HSMs.

To evaluate the hydrogen storage performance of LiBH<sub>4</sub>-catalyst composites, Ding et al.,<sup>[41a]</sup> for the first time, identified the primary factors influencing the dehydrogenation per-

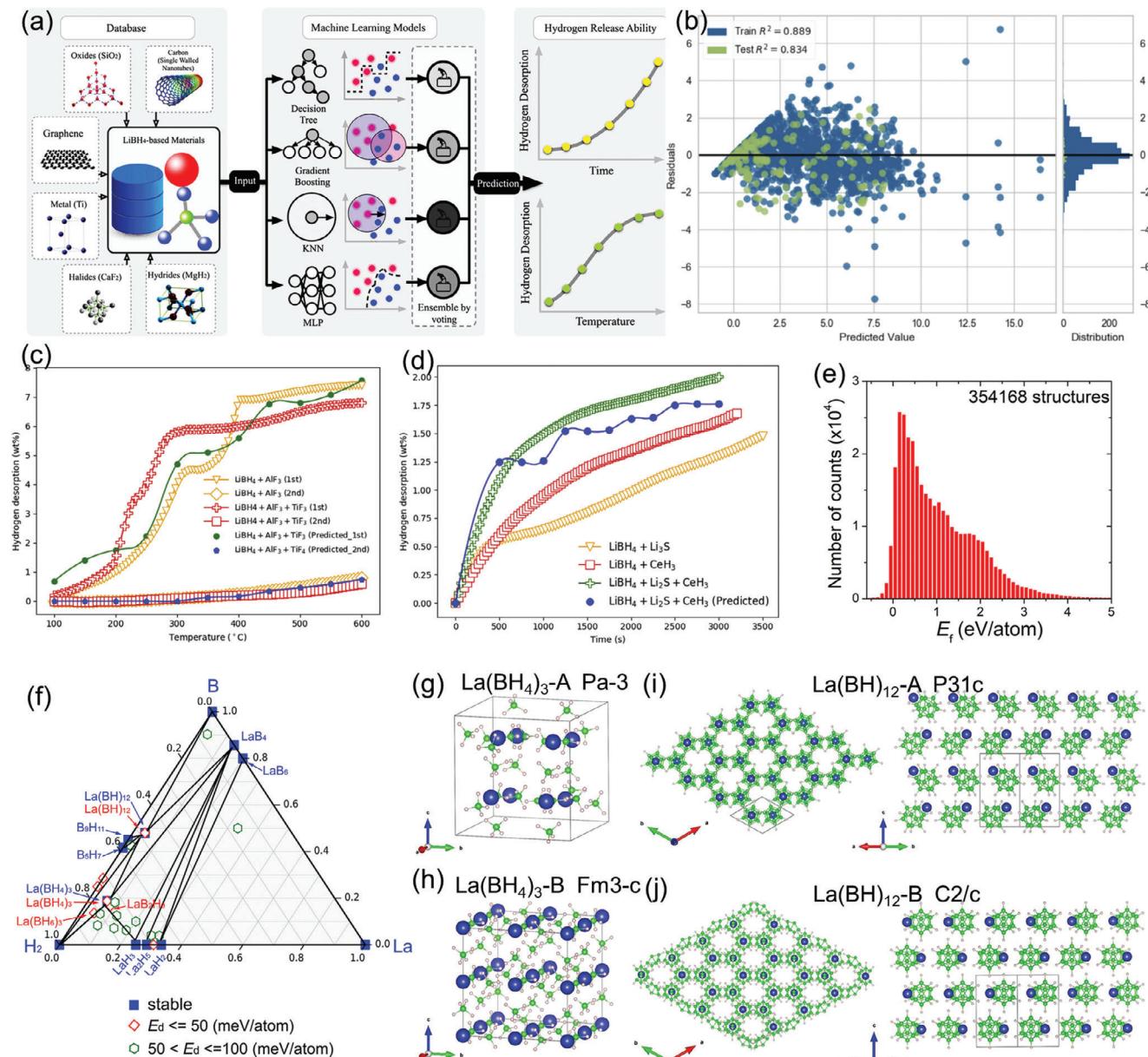
formance of LiBH<sub>4</sub> through the feature importance ranking of tree models, and these influential factors include dehydrogenation temperature, catalyst types, dehydrogenation atmosphere/catalyst content, etc. The ensemble of ensemble learning (EOE) model constructed with DT, GBDT, KNN and MLP base models demonstrates the lowest MSE of 1.144 and highest R<sup>2</sup> of 0.888 in predicting the doping of single-component catalyst. Moreover, the EOE model exhibits superior generalization capability in predicting both unknown single-component data and the performance of bi-component catalysts, as illustrated in Figure 13a–c. It is noteworthy that although TiF<sub>3</sub> addition enhances the dehydrogenation performance of LiBH<sub>4</sub>+AlF<sub>3</sub> mixture in the first round, it almost loses its dehydrogenation ability in the second round (~0.1 wt% H<sub>2</sub>), and this anomalous behavior is also accurately identified by the EOE black-box model, further confirming its predictive reliability (Figure 13d).

Despite extensive research, the composition-structure-property space of M-B-H systems remains largely uncharted owing to the immense variety of potential combinations of compositions and crystal structures involving three or more possible chemical elements. Notably, there have been no systematic investigations into the structure-property relationship of transition or RE metal *closoborates* and their related families. To delve into the realm of novel and energetically advantageous ternary metal borohydride compounds, Cheng et al.<sup>[144]</sup> identified 30 low-energy La-B-H ternary metal hydrides from 354 168 hypothetical materials with formation energy lower than 100 meV/atom compared to the convex hull by using an iteratively trained CGCNN model (MAE = 0.063 eV/atom), as exhibited in Figure 13e–f. It is noteworthy that two La(BH<sub>4</sub>)<sub>3</sub> phases and two La(BH)<sub>12</sub> phases exhibit either stability or formation energy of within 10 meV/atom over the convex hull (Figure 13g–j). Centering on these four La-B-H ternary compounds with the lowest energy, the impact of substitutions by elements from Groups 1, 2, 3, 13, and 14 was further explored. The findings confirm that the size of cation plays a pivotal role in shaping the structural attributes of metal *closododecaborates* and thus influencing the stability of ternary metal borohydrides.

Given the intricate de-/hydrogenation mechanisms of complex HSMs, there is a scarcity of relevant ML studies. Looking ahead, future related ML research will not only encompass systematic studies in the aspects of composite modification, catalytic modification, and nano-confinement modification, but will also involve the exploration of complex interatomic interactions using machine-learning interatomic potential functions. These endeavors will facilitate the theoretical identification of crucial factors that hinder the sluggish and complicate de-/hydrogenation processes, thus enabling high-throughput theoretical design of novel complex HSMs.

### 3.7. Solid-State Hydrogen Storage Devices

The significant thermal effect accompanying the de-/hydrogenation process of HSMs can severely hinder the kinetic rate and efficiency of de-/hydrogenation, and how to achieve efficient hydrogen storage as well as heat and mass transfer in metal hydride bed is one of the urgent engineering problems to be solved.<sup>[145]</sup> Currently, ML research on solid-state



**Figure 13.** a) Schematic diagram of ML process for the dehydrogenation ability of  $\text{LiBH}_4$ -based composites; b) Prediction results and c) residuals of dehydrogenation ability of  $\text{LiBH}_4$  with dual-component catalysts; d) First and second dehydrogenation curves of  $\text{LiBH}_4 + \text{AlF}_3$  and  $\text{LiBH}_4 + \text{AlF}_3 + \text{TiF}_3$ .<sup>[41a]</sup> a-d) Reproduced with permission.<sup>[41a]</sup> Copyright 2019, Elsevier Ltd. e) Frequency counts of formation energy of hypothetical ternary borohydride with CGCNN model; f) Phase diagram of La-B-H ternary borohydride; g-j) Geometrical structure of two  $\text{La}(\text{BH}_4)_3$  species and two  $\text{La}(\text{BH})_{12}$  species that are energetically stable.<sup>[144]</sup> e-j) Reproduced with permission.<sup>[144]</sup> Copyright 2024, American Chemical Society.

hydrogen storage devices focuses on Ti-based and RE-based HSMs that are capable of de-/hydrogenation under moderate conditions.

Prediction of the hydrogenation characteristics and hydrogenation state of solid-state hydrogen storage devices is of paramount importance for their practical implementation. Gao et al.<sup>[146]</sup> discovered that the hydrogenation curves of solid-state hydrogen storage devices based on  $\text{AB}_2$ -type HSMs exhibit an S-shape pattern with variation in temperature and rate, which can be effectively described by the Richards model. Subsequently,

the parameters of Richards model were fitted using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. Consequently, the results imply the exceptional predictive performance and robust fitting correlation ( $R^2 \geq 0.96$ ) obtained from the model based on linear regression. Therefore, this research offers an efficient methodology for predicting and evaluating the hydrogenation state of metal hydride-based hydrogen storage devices.

For hydrogen storage devices assembled with rare earth-based materials, ML studies include hydride bed design, mass and heat transfer rates and efficiencies, etc. Peng et al.<sup>[147]</sup> utilized

ML methods such as ensemble model and DT model to establish the relationship between hydrogenation duration and the dimensionless geometric parameters of fins (thickness, height, and width) and hydrogen storage capacity. Furthermore, they employed GA algorithm to obtain the optimal dimensionless fin configuration based on the optimal hydrogenation duration. As a result, the hydrogen storage capacity of the optimized metal hydride bed increased by 12.8% when the hydrogenation duration was 400 s (Figure 14a–c). Similarly, to establish the correlation between the hydrogen absorption and desorption efficiency ( $\eta_a$ ,  $\eta_d$ ) of LaNi<sub>5</sub> tubular reaction beds and the aspect ratio (AR) as well as the thermal conductivity ( $k_t$ ) of the chamber, Wang et al.<sup>[148]</sup> constructed a high-precision radial basis neural network (RBNN) model, which ultimately achieved a maximum relative error of less than 8.0% and 6.6% for hydrogen absorption and desorption efficiency of tubular reactor, respectively (as shown in Figure 14d–g). To ascertain the parameters of a commercial AB<sub>5</sub>-based hydride tank's mathematical model, Suarez et al.<sup>[149]</sup> introduced an identification approach applying evolutionary optimization algorithms such as GA, PSO, and DE (Differential Evolution, an algorithm that updates solutions via geometric manipulation of a subset of random individuals.) to assess the physical equations governing mass and heat transfer processes (Figure 14h,i). The findings reveal that GA is superior due to its precision and efficiency, while PSO is nearly as accurate. Meanwhile, DE stands out for its straightforward implementation and robustness against initial conditions. The simulation results of the hydrogen storage system, using the parameters derived from the algorithm, exhibit a remarkable congruence with experimental data, featuring a negligible maximum deviation of merely 2.1%. Tiwari et al.<sup>[41b]</sup> utilized a simple yet highly precise DT model to predict the de-/hydrogenation kinetic rates as well as the metal hydride bed temperature under varying parameters (time, hydrogen feeding pressure, flow rate of heat transfer oil, inlet temperature, etc.). Afterward, optimizations were successfully conducted towards hydrogen storage, space/water heating, and fuel cell integration systems with ML model and brute force algorithm. On a similar note, Zhao et al.<sup>[150]</sup> investigated the enhancement of heat and mass transfer efficiency in a LaNi<sub>5</sub>-based hydride reactor using graphene oxide nanofluids and ML models. Specifically, they applied gene expression programming (GEP), an evolutionary ML technique, to construct high-precision hydrogenation prediction models in both temperature-based and pressure-based manner (temperature-based model:  $R^2 = 0.9798$ , MAE = 0.046; pressure-based model:  $R^2 = 0.9946$ , MAE = 0.00267), which demonstrated the GEP technique can accurately capture the effect of operating parameters on the hydrogenation performance, providing a strong support for efficient hydrogen storage and thermal management strategies.

ML has achieved substantial progress in predicting and evaluating de-/hydrogenation characteristics and device design for solid-state hydrogen storage devices. However, the data sources currently utilized in related ML studies are primarily from basic HSMs and from numerical simulations. Consequently, future research should prioritize the exploration of optimized alloys for practical applications, large-scale hydrogen storage devices, and hydrogen storage bed design (including factors such as binder, thermal conductivity, and porosity).

## 4. Challenges of Machine Learning in Hydrogen Storage Materials

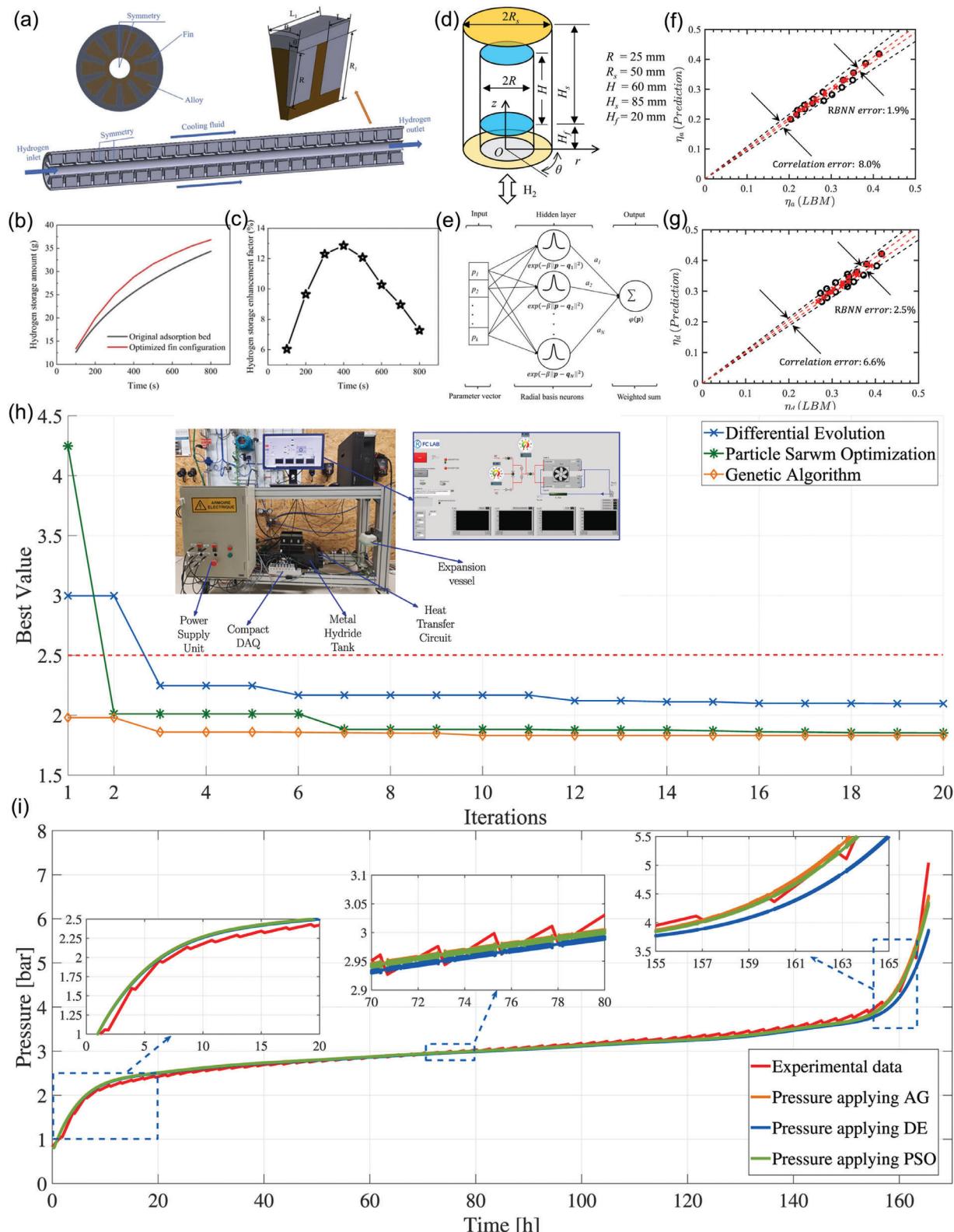
As previously mentioned, ML has not received extensive research focus in the realm of solid-state HSMs, and the primary challenges are summarized as follows: poor data quality and availability, and the balance between model interpretability and accuracy. These challenges will be described separately, together with practical solution ideas.

### 4.1. Poor Data Quality and Availability

Performance data for HSMs generally necessitate the acquisition from experiments, which are time-consuming and costly, resulting in a scarcity of high-quality datasets. Although the HydPARK dataset contains the majority of HSM systems, it suffers from several data quality problems (duplicates, missing, unreliable, etc.). Several derivative versions were parsed and cleaned up from the original HydPARK database by removing duplicate entries, removing entries missing critical data, and fixing various entries in the data with obvious errors, and updating the work with the new reports under version control.<sup>[26,27c,28,128,141]</sup> However, the size and quality of HydPARK-related datasets for specific material system are still unfavorable for high-quality ML. The thermodynamic hydrogen storage performance is first and foremost closely related to intrinsic factors such as elemental composition, crystal structure, grain size, atomic site preferences and ordering, stoichiometric ratios, and inherent hydrogenation capabilities.<sup>[152]</sup> However, the deviations in the experimental results of solid-state HSM of the same composition and structure are primarily attributed to external factors. These include material preparation/modification methods, mass of materials under test, de-/hydrogenation conditions (e.g., temperature, pressure, etc.), instrument factors for property determination (e.g., specific volumes, sensor sensitivity, hydrogen purity, etc.), and the design of hydrogen storage devices (e.g., shape, fin distribution, hydride bed composition, thermal conductivity agents, etc.).<sup>[153]</sup> Consequently, data acquired from different laboratories or under different conditions may demonstrate inconsistencies, which may compromise the accuracy and generalization capabilities of ML models.

To remedy the prevailing dilemma of data quality and availability for solid-state HSMs, feasible solutions include constructing a shared database of HSMs, data mining, and ML techniques that are effective even with small datasets. To build a shared database of solid-state HSMs, it is necessary to encourage research institutions/enterprises to share experimental data as well as develop a unified data standard and format.<sup>[26,29,141]</sup>

In addition, data mining based on natural language processing (NLP) and large language modeling (LLM) can address data limitations. Huang et al.<sup>[29]</sup> developed a Python-based tool that facilitates automatic downloading, data analysis and extraction, and data management of HSM-based papers. This tool utilizes NLP to intelligently extract crucial data such as chemical formulas, de-/hydrogenation temperature, and pressures from relevant papers. The extracted information pertains to Mg-based HSMs, Al-based HSMs, B-based HSMs, N-based HSMs, and carbon adsorbent materials, encompassing over 2000 data entries. Furthermore, they extracted 6468 hydrogen-containing



**Figure 14.** a) Schematic diagram of coaxial cylindrical reactor; b,c) The optimized hydrogen storage amount and enhancement factor with the function of duration time.<sup>[147]</sup> a–c) Reproduced with permission.<sup>[147]</sup> Copyright 2024, Elsevier Ltd. Schematic diagram of d)  $\text{LaNi}_5$ -based metal hydride bed and e) RBNN model; Parity plots of efficiency for f) hydrogenation and g) dehydrogenation process.<sup>[148]</sup> d–g) Reproduced with permission.<sup>[148]</sup> Copyright 2021, Elsevier Ltd. h) Optimization of best value with iterations; i) Pressure curve with time for different evolutionary optimization algorithms (h) insets with test desk and interface for the control software).<sup>[149]</sup> h,i) Reproduced with permission.<sup>[149]</sup> Copyright 2023, Elsevier Ltd.

materials from the existing material genome database (e.g., MP, OQMD, AFLOWLIB, etc.), involving seven crystal systems, 187 space groups, and 8118 crystal structures, and obtained their compositions, valences, bond lengths, and other physicochemical information.<sup>[29]</sup> By learning the bonding strength between H and other elements in different crystal field environments, it is helpful to predict and design the de-/hydrogenation performance of HSMs under complex compositions and processes.

Faced with the current situation of the scarce amount of sample data for solid-state HSMs, it is particularly important to adopt strategies such as data augmentation, active learning, transfer learning, unsupervised, and semi-supervised learning techniques for effective model training with limited data.<sup>[12b,89a,154]</sup> As for data augmentation, generative adversarial networks (GAN) and variational autoencoder (VAE) are typical representatives,<sup>[89b,155]</sup> which can generate substantial amounts of high-quality synthetic data to support the training of more complicated and powerful ML models. GAN draws on an adversarial training process between the generator and the discriminator, urging the generator to generate fake data with a distribution closely resembling that of the real data, while the discriminator is responsible for distinguishing between the real and fake data. Both parties improve through competition, ultimately achieving a dynamic equilibrium. For example, Ohno et al.<sup>[156]</sup> uses a GAN-based model for training data augmentation of material ionic conductivity for density estimation and sample generation. Besides, VAE is a deep learning architecture based on a probabilistic generative model. It utilizes an encoder to transform data into a latent space distribution, from which the decoder subsequently samples to create novel samples. VAE is frequently paired with GAN to bolster its performance and effectiveness. For example, Han et al.<sup>[157]</sup> adopted an improved HP-VAE-GAN model for generating microscopic images of materials for training set expansion. Allen et al.<sup>[158]</sup> generated 2D electrochemical impedance parameters for microbial corrosion using GAN and VAE, and the accuracy of generated data reaches over 83%, satisfying the requirements for subsequent ML training. As previously mentioned, despite limited in-depth research on data augmentation for solid-state HSMs, these advanced techniques can generate high-quality targets for de-/hydrogenation performance. The synthetic data closely resemble real data, enhancing model generalization and improving prediction accuracy for de-/hydrogenation behaviors of solid-state HSMs.

Active learning is an efficient method to optimize the training set construction strategy. By intelligently selecting samples for iterative learning, high-performance models are possible on a small dataset by utilizing only a small amount of sample data that is most critical for model performance improvement.<sup>[9e,159]</sup> This approach not only improves the data utilization efficiency, but also accelerates the model training process. Chen et al.<sup>[160]</sup> introduced an active learning-based data screening (ALDS) framework that effectively filters out unreliable external data, enabling the development of a highly precise model with minimal reliance on internal experimental data. Accordingly, training error of ALDS was reduced to 1.3% of the previous level, facilitating high-accuracy predictions for the negative expansion coefficient in the anti-perovskite structure. Therefore, by employing active learning techniques, it is possible to assemble a compact yet highly curated training set for solid-state HSMs. This, in turn,

allows the model to attain high-performance benchmarks with just a few iterations, which is of utmost importance for accelerating the development and refinement of models in the realm of solid-state HSMs research.

Considering the similarity in the aspects of hydrogen storage mechanisms, transfer learning is also a reliable choice. In detail, transfer learning is the ability to accelerate model adaptation and generalization to new domains with pre-trained models in related domains or similar tasks as a starting point, and quickly migrated to new materials in scenarios through fine-tuning techniques.<sup>[12b,30b,161]</sup> For instance, Batalovic et al.<sup>[140]</sup> designed a high-accuracy model for predicting hydride formation enthalpy of Mg-based HSMs, utilizing the initial crystal structure of intermetallic compounds and utilizing transfer learning with the Material Graph Network (MEGNet) as the foundation. Zhu et al.<sup>[162]</sup> achieved accurate bandgap predictions for diamond-structured chalcogenide thermoelectric materials by integrating 32 instances of material data into the initial 87 samples through a transfer learning algorithm. This approach exploits potential knowledge within the relevant dataset to facilitate new learning tasks, surpassing the performance of non-transfer models trained solely on the combined data. Considering the various forms of hydrogen storage, there could be underlying connections among interstitial HSMs, including Ti-based, RE-based, and solid solution HSMs. Given that hydrogen atoms in Mg-based and complex HSMs are bound through stable chemical bonds, they share comparable thermodynamic challenges in de-/hydrogenation processes. Consequently, by employing transfer learning techniques, it is feasible to develop high-precision ML models tailored to specific solid-state HSMs, using existing optimized models as a foundation.

What's more, in the case of limited data labeling resources, unsupervised learning methods can be used to explore the intrinsic structure of the data. Meanwhile, semi-supervised learning techniques (e.g., self-training, co-training, etc.) assist a small amount of labeled data for model training through unlabeled data to improve the generalization ability of the model, which is especially suitable for the case of insufficient data or high-cost of labeling.<sup>[163]</sup> For example, Chen et al.<sup>[164]</sup> proposed an innovative semi-supervised learning framework for the automatic segmentation of aluminum alloy metallographic images. By using a small number of labeled images and U-Net model, combined with a self-paced method and an improved pseudo-labeling selection strategy, this method achieves efficient segmentation of microstructures without relying heavily on manual labeling, and achieves excellent performance in experiments. By harnessing the potential value of unlabeled data and combining it directly with a modest amount of labeled data for model training, the generalization capability and predictive accuracy of the model can be substantially enhanced. These, these methodologies can effectively expedite the research and development process for solid-state HSMs.

#### 4.2. Balance Between Model Interpretability and High Precision

The conflict between interpretability and accuracy is a common situation for ML. High-precision models such as NN are accurate in their predictions, but their internal mechanisms

are complicated and intractable; whereas simple models with high interpretability, such as LR and DT, are mainly simple in structure, but cannot achieve the same prediction accuracy when dealing with complex hydrogen storage circumstances.<sup>[9e,165]</sup>

To realize the balance between interpretability and accuracy issues existing in ML, commonly adopted strategies include AutoML, hybrid models, rule extraction techniques, interpretability-enhanced model design, model transparency techniques, etc. A complex relationship generally exists between the composition-structure-properties of HSMs. Meanwhile, AutoML can automatically execute the entire ML process for the given data by means of random search, evolutionary optimization, meta-learning, etc., minimizing or eliminating manual intervention in ML process.<sup>[166]</sup> Up to date, AutoML is widely prevalent across various fields, including materials science and engineering, healthcare, surgery, economic task, satellite photo recognition, etc.<sup>[167]</sup> To effectively describe the concrete compressive strength, Sun et al.<sup>[168]</sup> utilized the AutoML model to automatically select models and adjust hyperparameters. The results demonstrate that AutoML-SHAP collaborative model can be automatically constructed within 3 minutes, outperforming other ML models without manual intervention. Since AutoML minimizes or eliminates the need for manual intervention throughout the entire ML process, it represents a highly attractive option for the nascent field of solid-state HSMs.

As for the hybrid model strategy, it typically involves using highly accurate but non-interpretable models for feature extraction, followed by inputting the extracted features into easily interpretable models that enhance the interpretability of the model while maintaining the prediction accuracy.<sup>[169]</sup> One of the representative examples is handwritten digital recognition of MNIST dataset. Ahlawat et al.<sup>[170]</sup> employed the CNN model to extract prominent features from the diverse and significantly distorted handwritten digit images, and then applied SVM as the classifier to categorize those extracted features. Consequently, this innovative CNN-SVM hybrid model significantly boosts the recognition performance and yields an impressive accuracy rate of 99.28%. In summary, hybrid models stand out for their capability to simultaneously boost both model interpretability and predictive accuracy, rendering them an outstanding option for predicting the de-/hydrogenation performance based on the geometrical structure of solid-state HSMs.

As a comparison, rule extraction techniques can convert the decision-making process of a complex model into a series of easy-to-understand if-then-else rules, which not only facilitate human beings to grasp the decision-making logic, but also allow approximating or substituting the original model with a set of rules in some cases.<sup>[9e,171]</sup> Bologna et al.<sup>[171c]</sup> achieved a remarkable feat by successfully extracting if-then-else rules from the linear combinations of dimension incremental multi-layer perceptron, ensuring a perfect 100% matching with the NN responses and exhibiting superior average accuracy over rules extracted from C4.5 DTs. Hence, rule extraction techniques, by offering deep insights into the decision-making processes of ML models, enhance our understanding of the intricate composition-structure-property relationships within solid-state HSMs. Consequently, they hold promise as one of the effective avenues for exploring the mechanisms underlying solid-state HSMs in the future.

In addition, interpretability-enhanced model design, such as linear models with sparse weights, shallow neural networks, and tree-based ensemble methods is also a favorable solution to reach a delicate trade-off between the interpretability and accuracy.<sup>[41a]</sup> To realize the prediction of bandgap, formation energy as well as the electronic configuration states of several functional materials/structures, Allen et al.<sup>[172]</sup> attained accuracy comparable to kernel methods and NN by employing linear combinations of nonlinear basis functions, enabling enhanced interpretability of the model's insights directly from the linear coefficients and functional forms. Therefore, reasonable model design holds the potential to achieve a balance between interpretability and model accuracy, representing one of the research directions in solid-state HSMs.

For model transparency techniques, it is possible to reveal how the model makes decisions derived from the inputting data with the assistance of feature importance assessment of tree-based ML models, global interpretation methods (e.g., SHAP), local interpretation methods (e.g., LIME), etc.<sup>[11d,40b,128,139,168,173]</sup> For example, Lin et al.<sup>[97]</sup> based on the feature importance ranking derived from the high-precision GBDT model, providing direct evidence of the significant contribution of the bandgap on the polycrystalline dielectric constant. Based on SHAP, Zhou et al.<sup>[11d]</sup> discovered a notable negative correlation between the Fe content and hydrogen storage capacity of C14-type HSMs. Their prediction revealed that the hydrogenation equilibrium pressure at room temperature reaches the GPa level. Consequently, while increasing the Fe content enhances the pressurization performance of the material, it comes at the expense of reduced hydrogen storage capacity. Thus, as described above, with suitable model transparency techniques, we can extract the design guidelines for high-performance solid-state HSMs.

## 5. Summary and Prospects

In summary, this review describes the datasets and feature descriptors suitable for solid-state HSMs as well as the selection of prevalent ML models. Afterward, the state-of-the-art ML research and application prospects for various solid-state HSMs are presented in detail, including mechanism mining, high-throughput design, and scale-up assembly application to deliver high hydrogen storage capacity with favorable de-/hydrogenation pathways and efficiency.

Despite demonstrating its certain potential in the research of solid-state HSMs, the application of ML in this field still faces numerous limitations. Firstly, ML models typically rely on a large amount of high-quality data, yet in HSMs research, high-quality experimental data and structural information are often scarce, which limits the accuracy and reliability of ML models. Secondly, current ML models for hydrogen storage performance prediction struggle to achieve a balance between accuracy (a direct representation of the model's predictive performance) and interpretability (how models make decisions), which is a common problem in ML research. In addition, ML investigations into some typical solid-state HSM systems have yet to be reported, and current ML research primarily focuses on isolated properties, lacking in-depth exploration of HSMs with comprehensive hydrogen storage performance. It should be noted that the application of ML in HSMs design is still in

its infancy, particularly in terms of integrating high-throughput design and inverse design techniques, with true breakthroughs yet to be achieved. Thus, based on the current research status quo, some research perspectives listed below hold promise for future investigations of high-performance solid-state HSMs.

- I. Regarding the dataset problem, on the one hand, it is necessary to build an on-the-fly shared HSM database and timely data mining. On the other hand, it is possible to generate an abundant high-quality data by means of data augmentation, or applied advanced ML approaches such as active learning, transfer learning, unsupervised and semi-supervised learning techniques, etc.
- II. To achieve a trade-off between accuracy and interpretability of hydrogen storage performance prediction models, it can be achieved by considering AutoML to minimize the human involvement for optimal model acquisition process, mixing of complex and easy-to-explain models, rule extraction, interpretability-enhanced model design and transparency of model interpretability.
- III. Systematic ML attempts to explore the hydrogen storage performance of Pd-based HSMs, TiFe-based HSMs, C15-type HSMs, RE-based HSMs, ZrCo-based HSMs,  $Zr_2M$ -based HSMs ( $M = Fe, Co, Ni$ ), complex HSMs, and multiphase hydrogen storage systems that received limited attention are urgently needed.
- IV. Integrating multi-objective optimization algorithms capable of effectively handling multiple objectives and constraints to effectively optimize the novel composition/structure with comprehensive hydrogen storage performance such as suitable de-/hydrogenation equilibrium pressure, superior kinetics, high hydrogen storage capacity, satisfactory cycling durability, etc.
- V. Machine learning interatomic potentials stands at the frontier of the intersection between materials science and ML, with the aim of simulating high-throughput discovery of materials and novel mechanisms at atomic scales. This area urgently requires research breakthroughs for the solid-state HSMs.
- VI. Machine learning-based inverse material design is an innovative approach that reverses the traditional forward design thinking from structure to performance, adopting instead a design approach from performance to structure. Accordingly, extraction of design guidelines (e.g., feature analysis from explainable ML models) and inverse material design (e.g., genetic algorithm (GA), particle swarm optimization (PSO), simulated annealing (SA), etc.) are urgently suggested.
- VII. ML for mass and heat transfer design to enhance the de-/hydrogenation efficiency of metal hydride beds/reactors, which is the critical concern for practical application of solid-state hydrogen storage/transport/use.

Thus, our review can serve as a valuable roadmap for researchers, guiding them to understand the current landscape of ML in solid-state HSMs, highlighting the key challenges, and outlining promising avenues for future investigations.

## Acknowledgements

P.Z. and Q.Z. contributed equally to this work. This work was financially supported by the National Natural Science Foundation of China (U24A2044, U20A20237, 52171223, 52071286), and the National Key Research and Development Program of China (2022YFB40043002).

## Conflict of Interest

The authors declare no conflict of interest.

## Keywords

high-throughput material design, hydrogen storage materials, machine learning, mechanism mining

Received: September 7, 2024  
Revised: November 10, 2024  
Published online:

- 
- [1] a) L. Schlapbach, A. Zuttel, *Nature* **2001**, *414*, 353; b) D. Q. Guan, B. W. Wang, J. G. Zhang, R. Shi, K. Jiao, L. C. Li, Y. Wang, B. A. Xie, Q. W. Zhang, J. Yu, Y. F. Zhu, Z. P. Shao, M. Ni, *Energ. Environ. Sci.* **2023**, *16*, 4926; c) J. O. Abe, A. P. I. Popoola, E. Ajenifaju, O. M. Popoola, *Int. J. Hydrot. Energy* **2019**, *44*, 15072; d) F. Dawood, M. Anda, G. M. Shafullah, *Int. J. Hydrot. Energy* **2020**, *45*, 3847; e) D. Lindley, *Nature* **2010**, *463*, 18.
  - [2] a) Y. Kojima, *Int. J. Hydrot. Energy* **2019**, *44*, 18179; b) S. Niaz, T. Manzoor, A. H. Pandith, *Renew. Sust. Energ. Rev.* **2015**, *50*, 457; c) M. R. Usman, *Renew. Sust. Energ. Rev.* **2022**, *167*, 112743; d) P. Zhou, J. W. Zhang, J. P. Bi, X. Z. Xiao, Z. M. Cao, L. J. Zhan, H. H. Shen, M. Lu, Z. N. Li, Y. Y. Zhao, L. Wang, M. Yan, L. X. Chen, *J. Mater. Chem. A* **2023**, *11*, 25963; e) M. D. Allendorf, V. Stavila, J. L. Snider, M. Witman, M. E. Bowden, K. Brooks, B. L. Tran, T. Autrey, *Nat. Chem.* **2022**, *14*, 1214; f) F. Otto, Y. Yang, H. Bei, E. P. George, *Acta Mater.* **2013**, *61*, 2628.
  - [3] J. T. Hu, J. J. Zhang, H. Y. Xiao, L. Xie, H. H. Shen, P. C. Li, J. W. Zhang, H. F. Gong, X. T. Zu, *Inorg. Chem.* **2020**, *59*, 9774.
  - [4] a) S. Dong, C. Q. Li, J. H. Wang, H. Liu, Z. Ding, Z. Y. Gao, W. J. Yang, W. Lv, L. Wei, Y. Wu, H. Li, *J. Mater. Chem. A* **2022**, *10*, 22363; b) S. Dong, C. Q. Li, E. F. Lv, J. H. Wang, H. Liu, Z. Y. Gao, W. Xiong, Z. Ding, W. J. Yang, H. Li, *J. Mater. Chem. A* **2022**, *10*, 19839.
  - [5] a) L. J. Zhan, P. P. Zhou, X. Z. Xiao, M. Lu, L. J. Jiang, Z. P. Li, M. Yan, L. X. Chen, *J. Energy Storage* **2024**, *89*, 111878; b) L. J. Zhan, Z. M. Cao, M. Y. Piao, X. Z. Xiao, P. P. Zhou, Y. P. Chen, Z. N. Li, L. J. Jiang, Z. P. Li, L. X. Chen, *Chem. Eng. J.* **2023**, *474*, 145654.
  - [6] K. Feldman, S. R. Agnew, *JOM* **2014**, *66*, 336.
  - [7] A. Agrawal, A. Choudhary, *Appl. Mater.* **2016**, *4*, 053208.
  - [8] a) Z. Y. Rao, P. Y. Tung, R. W. Xie, Y. Wei, H. B. Zhang, A. Ferrari, T. P. C. Klaver, F. Körmann, P. T. Sukumar, A. K. da Silva, Y. Chen, Z. M. Li, D. Ponge, J. Neugebauer, O. Gutfleisch, S. Bauer, D. Raabe, *Science* **2022**, *378*, 78; b) C. S. Wang, H. D. Fu, L. Jiang, D. Z. Xue, J. X. Xie, *NPJ Comput. Mater.* **2019**, *5*, 87; c) C. Wen, H. C. Shen, Y. W. Tian, G. Q. Lou, N. C. Wang, Y. J. Su, *Scr. Mater.* **2024**, *252*, 116240.
  - [9] a) S. Ishioka, A. Fujiwara, S. Nakanowatari, L. Takahashi, T. Taniike, K. Takahashi, *ACS Catal.* **2022**, *12*, 11541; b) X. F. Ma, Z. Li, L. E. K. Achenie, H. L. Xin, *J. Phys. Chem. Lett.* **2015**, *6*, 3528; c) W. B. Xu, M. Andersen, *Nat. Comput. Sci.* **2022**, *2*, 477; d) K. McCullough, T. Williams, K. Mingle, P. Jamshidi, J. Lauterbach, *Phys. Chem. Chem. Phys.* **2020**, *22*, 11174; e) Y. Liu, B. R. Guo, X. X. Zou, Y. J. Li, S. Q. Shi, *Energy Storage Mater.* **2020**, *31*, 434.

- [10] a) H. Zhou, B. X. Li, M. Yu, S. Li, G. Y. Fan, X. H. Ning, *Energy Storage Mater.* **2023**, *56*, 205; b) Z. B. Wei, X. F. Yang, Y. Li, H. W. He, W. H. Li, D. U. Sauer, *Energy Storage Mater.* **2023**, *56*, 62; c) M. Aykol, P. Herring, A. Anapolsky, *Nat. Rev. Mater.* **2020**, *5*, 725; d) M. F. Ng, J. Zhao, Q. Y. Yan, G. J. Conduit, Z. W. Seh, *Nat. Mach. Intell.* **2020**, *2*, 161.
- [11] a) G. Borboudakis, T. Stergiannakos, M. Frysali, E. Klontzas, I. Tsamardinos, G. E. Froudakis, *NPJ Comput. Mater.* **2017**, *3*, 40; b) M. I. M. Kusdhany, S. M. Lyth, *Carbon* **2021**, *179*, 190; c) S. Meduri, J. Nandanavanam, *Energy AI* **2023**, *12*, 100230; d) P. P. Zhou, X. Z. Xiao, X. Y. Zhu, Y. P. Chen, W. M. Lu, M. Y. Piao, Z. M. Cao, M. Lu, F. Fang, Z. N. Li, L. J. Jiang, L. X. Chen, *Energy Storage Mater.* **2023**, *63*, 102964; e) A. Verma, N. Wilson, K. Joshi, *Int. J. Hydrot. Energy* **2024**, *50*, 1518.
- [12] a) G. L. W. Hart, T. Mueller, C. Toher, S. Curtarolo, *Nat. Rev. Mater.* **2021**, *6*, 730; b) P. C. Xu, X. B. Ji, M. J. Li, W. C. Lu, *NPJ Comput. Mater.* **2023**, *9*, 42; c) Y. L. Liu, C. Niu, Z. Wang, Y. Gan, Y. Zhu, S. H. Sun, T. Shen, *J. Mater. Sci. Technol.* **2020**, *57*, 113; d) X. T. Zhong, B. Gallagher, S. S. Liu, B. Kaikhura, A. Hiszpanski, T. Y. J. Han, *NPJ Comput. Mater.* **2022**, *8*, 204; e) C. C. Gao, X. Min, M. H. Fang, T. Y. Tao, X. H. Zheng, Y. G. Liu, X. W. Wu, Z. H. Huang, *Adv. Funct. Mater.* **2022**, *32*, 2108044; f) G. Şenol, F. Selimfendigil, H. F. Öztürk, *Int. J. Hydrot. Energy* **2024**, *68*, 1178; g) G. C. Huang, F. Q. Huang, W. J. Dong, *Chem. Eng. J.* **2024**, *492*, 152294; h) G. Q. Wang, Z. M. Luo, G. D. H., M. Chen, Y. C. Dong, B. Lin, *Energ. Rev.* **2024**, *100106*.
- [13] a) S. M. Moosavi, K. M. Jablonka, B. Smit, *J. Am. Chem. Soc.* **2020**, *142*, 20273; b) M. I. Jordan, T. M. Mitchell, *Science* **2015**, *349*, 255; c) Y. Iwasaki, R. Sawada, V. Stanev, M. Ishida, A. Kirihara, Y. Omori, H. Someya, I. Takeuchi, E. Saitoh, S. Yorozu, *NPJ Comput. Mater.* **2019**, *5*, 103.
- [14] a) F. Noé, A. Tkatchenko, K. R. Müller, C. Clementi, *Annu. Rev. Phys. Chem.* **2020**, *71*, 361; b) S. J. Mooney, V. Pejaver, *Annual Review of Public Health* **2018**, *39*; c) D. Morgan, R. Jacobs, *Annu. Rev. Mater. Res.* **2020**, *50*, 71.
- [15] a) J. Y. Wang, P. C. Xu, X. B. Ji, M. J. Li, W. C. Lu, *Mater. Today Commun.* **2023**, *37*, 106910; b) Y. Liu, X. Zou, S. Ma, M. Avdeev, S. Shi, *Acta Mater.* **2022**, *238*, 118195; c) P. P. De Breuck, G. Hautier, G. M. Rignanese, *NPJ Comput. Mater.* **2021**, *7*, 83; d) A. S. Bundela, M. R. Rahul, *J. Alloys Compd.* **2022**, *908*, 164578.
- [16] a) S. Chong, S. Lee, B. Kim, J. Kim, *Coord. Chem. Rev.* **2020**, *423*; b) A. S. Ramesh, S. Vigneshwar, S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, W. Kim, *Fuel* **2023**, *337*, 126862; c) N. Wilson, A. Verma, P. R. Maharana, A. B. Sahoo, K. Joshi, *Int. J. Hydrot. Energy* **2024**, *90*, 460; d) G. M. Lu, M. Witman, S. Agarwal, V. Stavila, D. R. Trinkle, *Phys. Rev. Mater.* **2023**, *7*, 105402.
- [17] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *Apl. Mater.* **2013**, *1*, 011002.
- [18] S. Grazulis, D. Chateigner, R. T. Downs, A. F. T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, A. Le Bail, *J. Appl. Crystallogr.* **2009**, *42*, 726.
- [19] G. Bergerhoff, R. Hundt, R. Sievers, I. D. Brown, *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 66.
- [20] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, *JOM* **2013**, *65*, 1501.
- [21] C. Draxl, M. Scheffler, *MRS Bull.* **2018**, *43*, 676.
- [22] S. Curtarolo, W. Setyawan, S. D. Wang, J. K. Xue, K. S. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, O. Levy, *Comp. Mater. Sci.* **2012**, *58*, 227.
- [23] P. Villars, M. Berndt, K. Brandenburg, K. Cerzuai, J. Daams, F. Hulliger, T. Massalski, H. Okamoto, K. Osaki, A. Prince, H. Putz, S. Iwata, *J. Alloys Compd.* **2004**, *367*, 293.
- [24] P. J. Linstrom, W. G. Mallard, *J. Chem. Eng. Data* **2001**, *46*, 1059.
- [25] a) M. R. De Valladares, J. K. Jensen, in *4th Technol. Conv.*, Glasgow; b) HyMarc, **2022**.
- [26] HMARC, **2024**, **2019**.
- [27] a) A. Rahnama, G. Zepon, S. Sridhar, *Int. J. Hydrot. Energy* **2019**, *44*, 7337; b) A. Rahnama, G. Zepon, S. Sridhar, *Int. J. Hydrot. Energy* **2019**, *44*, 7345; c) M. Witman, S. L. Ling, D. M. Grant, G. S. Walker, S. Agarwal, V. Stavila, M. D. Allendorf, *J. Phys. Chem. Lett.* **2020**, *11*, 40; d) A. Rahnama, S. Sridhar, *Materialia* **2019**, *7*, 100366; e) J. Hattrick-Simpers, B. DeCost, *Mol. Syst. Des. Eng.* **2020**, *5*, 589; f) J. R. Hattrick-Simpers, K. Choudhary, C. Corgnale, *Mol. Syst. Des. Eng.* **2018**, *3*, 509.
- [28] M. Witman, G. Ek, S. L. Ling, J. Chames, S. Agarwal, J. Wong, M. D. Allendorf, M. Sahlberg, V. Stavila, *Chem. Mater.* **2021**, *33*, 4067.
- [29] P. Huang, D. Cai, H. Lin, J. Liu, Z. Li, B. Li, Y. Zou, H. Chu, L. Sun, F. Xu, *Sci. Sin. Chim.* **2022**, *52*, 1863.
- [30] a) C. M. Andreas, G. Sarah, *Introduction to Machine Learning with Python*, O'Reilly Media, Inc, Sebastopol, CA **2016**; b) M. D. Witman, S. L. Ling, M. Wadge, A. Bouzidi, N. Pineda-Romero, R. Clulow, G. Ek, J. M. Chames, E. J. Allendorf, S. Agarwal, M. D. Allendorf, G. S. Walker, D. M. Grant, M. Sahlberg, C. Zlotea, V. Stavila, *J. Mater. Chem. A* **2023**, *11*, 15878.
- [31] a) Y. Akkem, S. K. Biswas, A. Varanasi, *Eng. Appl. Artif. Intell.* **2023**, *120*, 105899; b) L. Zhang, J. Wen, Y. F. Li, J. L. Chen, Y. Y. Ye, Y. Y. Fu, W. Livingood, *Appl. Energy* **2021**, *285*, 116452; c) E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascárán, A. Fernández-Leal, *Artif. Intell. Rev.* **2023**, *56*, 3005.
- [32] S. Suwarno, G. Dicky, A. Suyuthi, M. Effendi, W. Witantyo, L. Noerochim, M. Ismail, *Int. J. Hydrot. Energy* **2022**, *47*, 11938.
- [33] S. Dangwal, Y. Ikeda, B. Grabowski, K. Edalati, *Chem. Eng. J.* **2024**, *493*, 152606.
- [34] X. He, J. D. Liu, C. Yang, G. Jiang, *Comp. Mater. Sci.* **2023**, *223*, 112111.
- [35] K. Batalovic, J. Radakovic, B. Kuzmanovic, M. M. Ilic, B. P. Mamula, *J. Energy Storage* **2023**, *68*, 107720.
- [36] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, *NPJ Comput. Mater.* **2016**, *2*, 16028.
- [37] a) T. Xie, J. C. Grossman, *Phys. Rev. Lett.* **2018**, *120*, 145301; b) C. W. Park, C. Wolverton, *Phys. Rev. Mater.* **2020**, *4*, 063801; c) P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. van Hoesel, H. Schopmans, T. Sommer, P. Friederich, *Commun. Mater.* **2022**, *3*, 93; d) Z. H. Dong, J. Feng, Y. J. Ji, Y. Y. Li, *J. Phys. Chem. A* **2023**, *127*, 5921.
- [38] a) Z. L. Lu, J. W. Wang, Y. F. Wu, X. M. Guo, W. Xiao, *Int. J. Hydrot. Energy* **2022**, *47*, 34583; b) Z. L. Lu, J. W. Wang, Y. F. Wu, X. M. Guo, T. R. Ma, W. Xiao, *Int. J. Hydrot. Energy* **2024**, *50*, 379.
- [39] a) F. Marques, M. Balcerzak, F. Winkelmann, G. Zepon, M. Felderhoff, *Energ. Environ. Sci.* **2021**, *14*, 5191; b) H. H. Shen, J. T. Hu, P. C. Li, G. Huang, J. W. Zhang, J. C. Zhang, Y. W. Mao, H. Y. Xiao, X. S. Zhou, X. T. Zu, X. G. Long, S. M. Peng, *J. Mater. Sci. Technol.* **2020**, *55*, 116; c) X. M. Wang, N. D. Tran, S. M. Zeng, C. Hou, Y. Chen, J. Ni, *NPJ Comput. Mater.* **2022**, *8*, 253.
- [40] a) W. O. Griffin, J. A. Darsey, *Int. J. Hydrot. Energy* **2013**, *38*, 11920; b) E. Halpren, X. Yao, Z. W. Chen, C. V. Singh, *Acta Mater.* **2024**, *270*, 119841.
- [41] a) Z. Ding, Z. Q. Chen, T. Y. Ma, C. T. Lu, W. H. Ma, L. Shaw, *Energy Storage Mater.* **2020**, *27*, 466; b) S. Tiwari, N. Gupta, S. Kumar, A. Kumar, P. Sharma, *J. Energy Storage* **2023**, *72*, 108522.
- [42] a) Z. R. Qiao, M. Welborn, A. Anandkumar, F. R. Manby, T. F. Miller, *J. Chem. Phys.* **2020**, *153*, 124111; b) K. T. Schütt, P. J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, K. R. Müller, *31st Annual Conference on Neural Information Processing Systems (NIPS)*, Long Beach, CA, **2017**.
- [43] H. X. Mai, T. C. Le, D. H. Chen, D. A. Winkler, R. A. Caruso, *Chem. Rev.* **2022**, *122*, 13478.

- [44] a) P. Friederich, F. Häse, J. Proppe, A. Aspuru-Guzik, *Nat. Mater.* **2021**, *20*, 750; b) A. Chen, X. Zhang, Z. Zhou, *Infomat* **2020**, *2*, 553.
- [45] a) P. Gil, H. Martins, F. Januário, *Appl. Soft Comput.* **2016**, *42*, 204; b) S. Jain, A. Saha, *Evol. Intell.* **2022**, *15*, 609.
- [46] a) H. AlSagri, M. Ykhlef, *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 628; b) H. J. Wang, Q. X. Liang, J. T. Hancock, T. M. Khoshgoftaar, *J. Big Data* **2024**, *11*, 44; c) Z. G. Wang, X. Xiao, S. Rajasekaran, *Big data Min. Anal.* **2020**, *3*, 208.
- [47] R. B. Wexler, J. M. P. Martirez, A. M. Rappe, *J. Am. Chem. Soc.* **2018**, *140*, 4678.
- [48] a) J. R. Wu, Y. G. Wang, *IEEE Trans. Emerg. Topics Comput.* **2023**, *7*, 513; b) D. H. Liu, H. Qian, G. Dai, Z. H. Zhang, *Pattern Recogn.* **2013**, *46*, 2531; c) X. J. Ding, F. Yang, S. Jin, J. Cao, *Comput. Biol. Med.* **2021**, *134*, 104505.
- [49] a) N. Wagner, J. M. Rondinelli, *Front. Mater.* **2016**, *3*, 28; b) W. Z. Li, M. J. Liu, *ACS Appl. Electron. Mater.* **2023**, *5*, 4523.
- [50] a) Q. Shao, C. J. Feng, J. Kang, *International Conference on Measurement, Instrumentation and Automation (ICMIA 2012)*, Guangzhou, China, **2012**, pp. 15–16; b) R. Hamamoto, K. Takasawa, H. Machino, K. Kobayashi, S. Takahashi, A. Bolatkan, N. Shinkai, A. Sakai, R. Aoyama, M. Yamada, K. Asada, M. Komatsu, K. Okamoto, H. Kameoka, S. Kaneko, *Brief. Bioinform.* **2022**, *23*, bbac246; c) J. Z. Gan, T. Liu, L. Li, J. L. Zhang, *Comput. J.* **2021**, *64*, 1080; d) Z. Z., Y. B. R., *Comput. Sci.* **2009**, *36*, 234.
- [51] a) W. Abdelbaki, N. Saad, K. Younes, *IEEE Access* **2024**, *12*, 92442; b) E. Obeid, H. Chaouk, R. Mezher, E. G. Hanna, O. Mouhtady, J. Halwani, K. Younes, *Gels* **2023**, *9*, 935.
- [52] a) S. Nanda, S. Ghosh, T. Thomas, *J. Power Sources* **2022**, *546*, 231975; b) T. Tiwari, S. Jalalian, C. Mendis, D. Eskin, *JOM* **2023**, *75*, 4526; c) C. Yang, Y. Kim, S. Ryu, G. X. Gu, *Mater. Des.* **2020**, *189*, 108509; d) S. Bhattacharya, A. Roy, *Comp. Mater. Sci.* **2024**, *231*, 112581.
- [53] a) J. Q. Fan, D. Wang, K. Z. Wang, Z. W. Zhu, *Ann. Stat.* **2019**, *47*, 3009; b) G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, T. Baker, *IEEE Access* **2020**, *8*, 54776; c) D. W. Jiang, J. L. Zhang, Z. Y. Wang, C. F. Feng, K. X. Jiao, R. Z. Xu, *JOM* **2020**, *72*, 3908.
- [54] a) J. Y. Yu, B. C. Pan, S. S. Yu, M. F. Leung, *Math. Biosci. Eng.* **2023**, *20*, 12486; b) L. L. Zong, X. C. Zhang, L. Zhao, H. Yu, Q. L. Zhao, *Neural Networks* **2017**, *88*, 74.
- [55] a) M. Kimura, *IEEE Access* **2021**, *9*, 129619; b) R. Silva, P. Melo-Pinto, *Artif. Intell. Agric.* **2023**, *7*, 58.
- [56] a) B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, *Phys. Rev. B* **2014**, *89*, 094104; b) C. Wen, Y. Zhang, C. X. Wang, D. Z. Xue, Y. Bai, S. Antonov, L. H. Dai, T. Lookman, Y. J. Su, *Acta Mater.* **2019**, *170*, 109; c) O. Isayev, C. Osse, C. Toher, E. Gossett, S. Curtarolo, A. Tropsha, *Nat. Commun.* **2017**, *8*, 15679; d) L. Himanen, A. Geurts, A. S. Foster, P. Rinke, *Adv. Sci.* **2019**, *6*, 1900808; e) J. P. Correa-Baena, K. Hippalgaonkar, J. van Duren, S. Jaffer, V. R. Chandrasekhar, V. Stevanovic, C. Wadia, S. Guha, T. Buonassisi, *Joule* **2018**, *2*, 1410.
- [57] a) Y. Liu, O. C. Esan, Z. F. Pan, L. An, *Energy and AI* **2021**, *3*, 100049; b) S. J. Ma, Y. X. Du, S. S. Wang, Y. J. Su, *Corros. Rev.* **2023**, *41*, 417; c) Z. Y. Zhang, A. M. Tehrani, A. O. Oliynyk, B. Day, J. Brugoch, *Adv. Mater.* **2021**, *33*, 2005112.
- [58] P. Wang, E. Fan, P. Wang, *Pattern Recogn. Lett.* **2021**, *141*, 61.
- [59] a) Imran, N. I., D. H. Kim, *Adv. Theory Simul.* **2022**, *5*, 2200517; b) A. Chen, J. F. Cai, Z. L. Wang, Y. Q. Han, S. M. Ye, J. J. Li, *J. Energy Chem.* **2023**, *78*, 268; c) N. Arirth, A. Urban, *Comp. Mater. Sci.* **2016**, *114*, 135.
- [60] Y. W. Chen, L. D. Zhou, S. W. Pei, Z. W. Yu, Y. Chen, X. Liu, J. X. Du, N. X. Xiong, *IEEE T. Syst. Man Cy-S* **2021**, *51*, 3939.
- [61] a) D. Sahin, S. Akleylek, E. Kilic, *IEEE Access* **2022**, *10*, 14246; b) V. R. Joseph, *Stat. Anal. Data Min.* **2022**, *15*, 531.
- [62] a) Y. J. Ran, X. G. Sun, *Comput. Appl. Softw.* **2013**, *30*, 6; b) H. Yildirim, M. R. Özkale, *Neural Comput. Appl.* **2021**, *33*, 10469.
- [63] G. T. Wang, P. Li, C. L. Su, *Inf. Control* **2011**, *40*, 497.
- [64] a) C. Wang, J. Chen, Y. J. Liu, *J. Electrochem. Soc.* **2023**, *170*, 120502; b) L. H. Guo, *Intell. Autom. Soft Co.* **2020**, *26*, 421.
- [65] a) J. Hu, Z. Cao, Y. Dan, C. Niu, X. Li, S. Qian, *J. South Chin. Uni. Technol.* **2019**, *47*, 48; b) R. A. S. Ferreira, S. F. H. Correia, L. S. Fu, P. Georgieva, M. Antunes, P. S. André, *Sci. Rep.* **2024**, *14*, 4160; c) B. V. Goodarzi, A. R. Bahramian, *J. Therm. Anal. Calorim.* **2022**, *147*, 6227.
- [66] P. L. Donti, J. Z. Kolter, *Ann. Rev. Environ. Resources* **2021**, *46*, 719.
- [67] a) U. Aich, S. Banerjee, *Appl. Math. Model.* **2014**, *38*, 2800; b) H. P. Zhang, X. C. Li, M. N. Amin, A. A. A. Al-Naghi, S. Ul Arifeen, F. Althoe, A. Ahmad, *Mater. Today Commun.* **2024**, *38*, 108543.
- [68] a) Z. Pan, Y. G. Zhou, L. Zhang, *ACS Appl. Mater. Interfaces* **2022**, *14*, 9933; b) Y. M. Peng, C. Unluer, *J. Clean.* **2024**, *434*, 140054; c) J. S. Chou, C. F. Tsai, A. D. Pham, Y. H. Lu, *Constr. Build. Mater.* **2014**, *73*, 771.
- [69] M. He, L. Zhang, *Comp. Mater. Sci.* **2021**, *196*, 110578.
- [70] M. Rahimi, M. H. Abbaspour-Fard, A. Rohani, *J. Clean.* **2021**, *329*, 129714.
- [71] T. Sutojo, S. Rustad, M. Akrom, A. Syukur, G. F. Shidik, H. K. Dipojono, *NPJ Mater. Degrad.* **2023**, *7*, 18.
- [72] a) Y. Sun, B. B. Hu, Y. T. Zhang, X. L. Song, J. Q. Feng, Y. Xu, H. Tao, D. Ergu, *J. Alloys Compd.* **2024**, *1003*, 175598; b) S. Kumar, H. Pradhan, N. Shah, M. R. Rahul, G. Phanikumar, *Scr. Mater.* **2023**, *234*, 115543.
- [73] a) R. Nia, M. Jalali, M. Houshmand, *Appl. Sci.-Basel* **2022**, *12*, 8021; b) X. P. Wu, F. Zhu, M. M. Zhou, M. M. S. Sabri, J. D. Huang, *Materials* **2022**, *15*, 4582; c) S. Sapkal, B. Kandasubramanian, P. Dixit, H. S. Panda, *Mater. Today Energy* **2023**, *37*, 101402.
- [74] I. D. Mienye, N. Jere, *IEEE Access* **2024**, *12*, 86716.
- [75] L. Chi, M. Wang, K. H. Liu, S. Lu, L. L. Kan, X. M. Xia, C. D. Huang, *Mater. Today Commun.* **2023**, *36*, 106470.
- [76] R. D. Párizs, D. Török, T. Ageyeva, J. G. Kovács, *Sensors* **2022**, *22*, 2704.
- [77] a) G. A. Landrum, H. Genin, *J. Solid State Chem.* **2003**, *176*, 587; b) H. Kurban, M. Kurban, *Chem. Phys.* **2021**, *546*, 111159; c) H. R. Zhang, X. Liu, G. J. Zhang, Y. Q. Zhu, S. Z. Li, Q. Qian, D. B. Dai, R. C. Che, T. Xu, *Comp. Mater. Sci.* **2023**, *228*, 112349.
- [78] a) S. H. Wang, H. S. Pillai, S. W. Wang, L. E. K. Achenie, H. L. Xin, *Nat. Commun.* **2021**, *12*, 5288; b) A. M. Tehrani, A. O. Oliynyk, M. Parry, Z. Rizvi, S. Couper, F. Lin, L. Miyagi, T. D. Sparks, J. Brugoch, *J. Am. Chem. Soc.* **2018**, *140*, 9844; c) X. Y. He, Y. X. Chen, S. Wang, G. Q. Zhang, *ACS Appl. Mater. Interfaces* **2024**, *16*, 24494.
- [79] C. Garbin, X. Q. Zhu, O. Marques, *Multimed. Tools Appl.* **2020**, *79*, 12777.
- [80] a) V. K. Ojha, A. Abraham, V. Snásel, *Eng. Appl. Artif. Intell.* **2017**, *60*, 97; b) T. Zhang, D. G. Zhang, H. R. Yan, J. N. Qiu, J. X. Gao, *Neurocomputing* **2021**, *420*, 98.
- [81] a) P. Zhang, Z. Y. Yin, Y. F. Jin, T. H. T. Chan, F. P. Gao, *Geosci. Front.* **2021**, *12*, 441; b) L. Wang, Y. Zeng, T. Chen, *Expert Syst. Appl.* **2015**, *42*, 855.
- [82] J. T. Huang, J. T. Xue, M. W. Li, J. Y. Chen, Y. Cheng, Z. H. Lai, J. Hu, F. Zhou, N. Qu, Y. Liu, J. C. Zhu, *Appl. Surf. Sci.* **2023**, *635*, 157757.
- [83] R. Y. Li, Q. Deng, D. Tian, D. Y. Zhu, B. Lin, *Crystals* **2021**, *11*, 818.
- [84] a) K. Cheng, X. D. Yang, X. P. Wu, *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5919510; b) X. Y. Zhu, Z. H. Zhang, Y. Mao, Y. Li, X. Huang, N. Gu, *Sci. China Mater.* **2020**, *63*, 2365; c) I. U. Khan, S. Afzal, J. W. Lee, *Sensors* **2022**, *22*, 323.
- [85] X. Fang, Z. Shi, *Comput. Eng.* **2018**, *44*, 218.
- [86] S. Back, J. Yoon, N. H. Tian, W. Zhong, K. Tran, Z. W. Ulissi, *J. Phys. Chem. Lett.* **2019**, *10*, 4401.

- [87] a) C. W. Coley, R. Barzilay, W. H. Green, T. S. Jaakkola, K. F. Jensen, *J. Chem. Inf. Model.* **2017**, *57*, 1757; b) X. Y. Lu, Z. Z. Xie, X. J. Wu, M. M. Li, W. Q. Cai, *Chem. Eng. Sci.* **2022**, *259*, 117813; c) Z. H. Wu, S. R. Pan, F. W. Chen, G. D. Long, C. Q. Zhang, P. S. Yu, *IEEE Trans. Neural Networks Learn. Syst.* **2021**, *32*, 4.
- [88] C. Chen, W. K. Ye, Y. X. Zuo, C. Zheng, S. P. Ong, *Chem. Mater.* **2019**, *31*, 3564.
- [89] a) Y. Zhang, C. Ling, *NPJ Comput. Mater.* **2018**, *4*, 25; b) L. J. Shen, Q. Qian, *Comp. Mater. Sci.* **2022**, *211*, 111475; c) K. M. Li, D. Persaud, K. Choudhary, B. Decost, M. Greenwood, J. Hattrick-Simpers, *Nat. Commun.* **2023**, *14*, 7283.
- [90] a) X. B. Dong, Z. W. Yu, W. M. Cao, Y. F. Shi, Q. L. Ma, *Front. Comput. Sci.* **2020**, *14*, 241; b) I. D. Miénye, Y. X. Sun, *IEEE Access* **2022**, *10*, 99129; c) L. H. Chen, B. Xu, J. Chen, K. Bi, C. J. Li, S. Y. Lu, G. S. Hu, Y. H. Lin, *J. Mater. Chem. C* **2020**, *8*, 13079; d) T. Dissanayake, Y. Rajapaksha, R. Ragel, I. Nawinne, *Sensors* **2019**, *19*, 4495.
- [91] M. Kaur, H. K. Gianey, D. Singh, M. Sabharwal, *Mod. Phys. Lett. B* **2019**, *33*, 1950022.
- [92] R. Kronberg, H. Lappalainen, K. Laasonen, *J. Phys. Chem. C* **2021**, *125*, 15918.
- [93] H. V. Thanh, S. E. Taremsari, B. Ranjbar, H. Mashhadimoslem, E. Rahimi, M. Rahimi, A. Elkamel, *Energies* **2023**, *16*, 2348.
- [94] a) X. L. Zhang, R. Fang, *Appl. Res. Comput.* **2009**, *26*, 77; b) L. Li, C. Y. Wang, W. Li, J. B. Chen, *Neurocomputing* **2018**, *275*, 1725.
- [95] a) L. F. Ma, H. M. Xiao, J. W. Tao, T. Y. Zheng, H. Q. Zhang, *Open Geosci* **2022**, *14*, 629; b) Y. Agrawal, M. Kumar, S. Ananthakrishnan, G. Kumarapuram, *Water Resources Management* **2022**, *36*, 1025.
- [96] H. C. Chen, W. C. Shan, H. Y. Liao, Y. X. He, T. Zhang, P. C. Pei, C. H. Deng, J. R. Chen, *Int. J. Hydrol. Energy* **2021**, *46*, 34399.
- [97] X. P. Yang, C. Yuan, S. R. He, D. Jiang, B. Cao, S. Wang, *Fuel* **2023**, *331*, 125718.
- [98] a) H. Y. Liu, J. C. Cheng, H. Z. Dong, J. G. Feng, B. L. Pang, Z. Y. Tian, S. Ma, F. J. Xia, C. K. Zhang, L. F. Dong, *Comp. Mater. Sci.* **2020**, *177*, 109614; b) Z. Zhang, H. Y. Gao, Z. W. Zhang, P. Chen, Y. He, L. Guo, Z. G. Ding, J. Chen, C. T. Wang, Y. He, *Mater. Today Commun.* **2024**, *39*, 109070.
- [99] a) M. Koopalipoor, P. G. Asteris, A. S. Mohammed, D. E. Alexakis, A. Mamou, D. J. Armaghani, *Transp. Geotech.* **2022**, *34*, 100756; b) H. Rajadurai, U. D. Gandhi, *Neural Comput. Appl.* **2022**, *34*, 15387; c) T. Zhou, H. Jiao, *Educ. Psychol. Meas.* **2023**, *83*, 831.
- [100] T. Wang, K. F. Zhang, J. Thé, H. S. Yu, *Comp. Mater. Sci.* **2022**, *201*, 110899.
- [101] H. B. Li, X. Y. Li, P. T. Wang, Z. Zhang, K. Davey, Q. F. Shi, S. Z. Qiao, *J. Am. Chem. Soc.* **2024**, *146*, 22850.
- [102] a) A. Mishra, L. Kompella, L. M. Sanagavarapu, S. Varam, *Comp. Mater. Sci.* **2022**, *210*, 111025; b) R. Malhotra, *Appl. Soft Comput.* **2015**, *27*, 504; c) M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, *Nature* **2019**, *566*, 195; d) H. Y. Huang, R. Kueng, G. Torlai, V. V. Albert, J. Preskill, *Science* **2022**, *377*, 682.
- [103] C. Lu, Y. T. Shi, Y. Y. Chen, S. Q. Bao, L. X. Tang, *7th International Conference on Cloud Computing and Big Data (CCBD)*, Macau, China, **2016**.
- [104] a) H. W. Xie, X. L. Li, *Comput. Eng.* **2009**, *35*, 179; b) K. P. Sinaga, M. S. Yang, *IEEE Access* **2020**, *8*, 80716.
- [105] a) Y. Liu, H. Zhang, *Comput. Eng. Appl.* **2012**, *48*, 142; b) Y. Zhang, P. P. Fu, Y. T. Zhang, *J. Comput. Appl.* **2013**, *33*, 2801.
- [106] a) M. Unver, N. Erginel, *J. Intell. Fuzzy Syst.* **2020**, *39*, 6099; b) M. M. Liu, T. Y. Wang, Q. Y. Zhang, C. B. Pan, S. H. Liu, Y. M. Chen, D. Lin, S. Y. Feng, *Anal. Methods* **2024**, *16*, 846.
- [107] a) V. L. Deringer, M. A. Caro, G. Csányi, *Adv. Mater.* **2019**, *31*, 1902765; b) C. D. Lv, X. Zhou, L. X. Zhong, C. S. Yan, M. Srinivasan, Z. W. Seh, C. T. Liu, H. G. Pan, S. Z. Li, Y. G. Wen, Q. Y. Yan, *Adv. Mater.* **2022**, *34*, 2101474; c) L. F. Zhang, D. Y. Lin, H. Wang, R. Car, *Phys. Rev. Mater.* **2019**, *3*, 023804.
- [108] a) K. W. Wan, J. X. He, X. H. Shi, *Adv. Mater.* **2024**, *36*, 2305758; b) X. H. Wan, Z. Y. Li, W. Yu, A. Y. Wang, X. Ke, H. L. Guo, J. H. Su, L. Li, Q. Z. Gui, S. P. Zhao, J. Robertson, Z. F. Zhang, Y. Z. Guo, *Adv. Mater.* **2023**, *202305192*; c) N. Nedik, R. Zubatyuk, M. Kulichenko, N. Lubbers, J. S. Smith, B. Nebgen, R. Messerly, Y. W. Li, A. I. Boldyrev, K. Barros, O. Isayev, S. Tretiak, *Nat. Rev. Chem.* **2022**, *6*, 653.
- [109] J. Behler, *Chem. Rev.* **2021**, *121*, 10037.
- [110] J. Behler, *J. Chem. Phys.* **2011**, *134*, 074106.
- [111] H. Wang, L. F. Zhang, J. Q. Han, *Comput. Phys. Commun.* **2018**, *228*, 178.
- [112] S. Batzner, A. Musaelian, L. X. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, B. Kozinsky, *Nat. Commun.* **2022**, *13*, 2453.
- [113] D. Guo, C. Li, K. Li, B. Shao, D. Chen, Y. Ma, J. Sun, X. Cao, W. Zeng, X. Chang, *Mater. Today Energy* **2021**, *20*, 100665.
- [114] F. Z. Dai, B. Wen, Y. J. Sun, H. M. Xiang, Y. C. Zhou, *J. Mater. Sci. Technol.* **2020**, *43*, 168.
- [115] a) R. van Mastrigt, M. Dijkstra, M. van Hecke, C. Coulais, *Phys. Rev. Lett.* **2022**, *129*, 198003; b) K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, *559*, 547; c) J. G. Greener, S. M. Kandathil, L. Moffat, D. T. Jones, *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 40.
- [116] a) J. W. Yeh, S. K. Chen, S. J. Lin, J. Y. Gan, T. S. Chin, T. T. Shun, C. H. Tsau, S. Y. Chang, *Adv. Eng. Mater.* **2004**, *6*, 299; b) S. Dangwal, K. Edalati, *Scr. Mater.* **2024**, *238*, 115774; c) X. Huang, X. Z. Xiao, X. C. Wang, Z. D. Yao, C. T. Wang, X. L. Fan, L. X. Chen, *Energy Storage Mater.* **2018**, *13*, 199; d) D. Dayan, M. H. Mintz, M. P. Dariel, *J. Less Common Met.* **1980**, *73*, 15; e) K. L. Lim, H. Kazemian, Z. Yaakob, W. R. W. Daud, *Chem. Eng. Technol.* **2010**, *33*, 213; f) X. Wang, W. Guo, Y. Z. Fu, *J. Mater. Chem. A* **2021**, *9*, 663.
- [117] a) C. E. Lundin, F. E. Lynch, C. B. Magee, *J. Less Common Met.* **1977**, *56*, 19; b) M. M. Nygard, G. Ek, D. Karlsson, M. H. Sorby, M. Sahlberg, B. C. Hauback, *Acta Mater.* **2019**, *175*, 121; c) A. Mohammadi, Y. Ikeda, P. Edalati, M. Mito, B. Grabowski, H. W. Li, K. Edalati, *Acta Mater.* **2022**, *236*, 118117; d) J. W. Ren, N. M. Musyoka, H. W. Langmi, M. Mathe, S. J. Liao, *Int. J. Hydrol. Energy* **2017**, *42*, 289; e) Y. J. Ma, Y. Ma, Q. S. Wang, S. Schweidler, M. Botros, T. T. Fu, H. Hahn, T. Brezesinski, B. Breitung, *Energ. Environ. Sci.* **2021**, *14*, 2883; f) N. A. A. Rusman, M. Dahari, *Int. J. Hydrol. Energy* **2016**, *41*, 12108.
- [118] S. Nations, T. Nandi, A. Ramazani, S. N. Wang, Y. H. Duan, *J. Energy Storage* **2023**, *70*, 107980.
- [119] a) P. P. Zhou, Z. M. Cao, X. Z. Xiao, Z. F. Jiang, L. J. Zhan, Z. N. Li, L. J. Jiang, L. X. Chen, *Int. J. Hydrol. Energy* **2022**, *47*, 1710; b) P. P. Zhou, Z. M. Cao, X. Z. Xiao, L. J. Zhan, S. Q. Li, Z. N. A. Li, L. J. Jiang, L. X. Chen, *J. Alloys Compd.* **2021**, *875*, 160035; c) V. A. Yartys, M. V. Lototskyy, *J. Alloys Compd.* **2022**, *916*, 165219; d) M. Piao, X. Xiao, L. Zhan, Z. Cao, P. Zhou, J. Qi, M. Lu, Z. Li, L. Jiang, F. Fang, L. Chen, *Int. J. Hydrol. Energy* **2024**, *50*, 1358; e) X. M. Guo, S. M. Wang, X. P. Liu, Z. N. Li, F. Lue, J. Mi, L. Hao, L. J. Jiang, *Rare Met.* **2011**, *30*, 227; f) W. F. Qiao, W. Q. Liu, D. M. Yin, N. Ding, S. L. Zhao, H. X. Xiu, C. Liu, Y. Wang, M. He, C. L. Wang, L. M. Wang, Y. Cheng, *Chem. Eng. J.* **2024**, *495*, 153489; g) J. B. Ponsoni, V. Aranda, T. d. S. Nascimento, R. B. Strozi, W. J. Botta, G. Zepon, *Acta Mater.* **2022**, *240*, 118317.
- [120] M. Gheytanzadeh, F. Rajabhasani, A. Baghban, S. Habibzadeh, O. Abida, A. Esmaeili, M. T. Munir, *Sci. Rep.* **2022**, *12*, 21902.
- [121] S. Maghsoudy, P. Zakerabbasi, A. Baghban, A. Esmaeili, S. Habibzadeh, *Sci. Rep.* **2024**, *14*, 1503.
- [122] J. M. Kim, T. Ha, J. Lee, Y.-S. Lee, J.-H. Shim, *Met. Mater. Int.* **2023**, *29*, 861.
- [123] a) S. N. Klyamkin, M. T. Hagström, E. V. Mescheryakova, P. D. Lund, *J. Mater. Sci.* **2000**, *35*, 133; b) J. J. Liu, H. H. Cheng, S. M. Han, H. F. Liu, J. Huot, *Energy* **2020**, *192*, 116617; c) J. J. Liu, X. Y. Chen, J. Xu, S. Zhu, H. H. Cheng, G. Yang, X. B. Han, L. Zhang, Y. Li, S. M. Han,

- Chem. Eng. J.* **2021**, *418*, 129395; d) B. Liao, Y. Q. Lei, L. X. Chen, G. L. Lu, H. G. Pan, Q. D. Wang, *J. Alloys Compd.* **2004**, *376*, 186; e) S. Han, X. Zhang, S. Shi, H. Tanaka, N. Kuriyama, N. Taoka, K. Aihara, Q. Xu, *J. Alloys Compd.* **2007**, *446*, 208; f) F. Schwepppe, M. Martin, E. Fromm, *J. Alloys Compd.* **1997**, *253*, 511.
- [124] Z. Shuang, L. Qin, C. Ning, M. Li, Y. Wen, *J. Alloys Compd.* **1999**, *287*, 57.
- [125] S. Zhao, Chongqing University, Chongqing **2012**.
- [126] a) Z. J. Cao, L. Z. Ouyang, H. Wang, J. W. Liu, L. X. Sun, M. Felderhoff, M. Zhu, *Int. J. Hydrot. Energy* **2016**, *41*, 11242; b) J. J. Liu, J. Xu, S. Sleiman, X. Y. Chen, S. Zhu, H. H. Cheng, J. Huot, *Int. J. Hydrot. Energy* **2021**, *46*, 28709.
- [127] Z. Lu, J. Wang, Y. Wu, X. Guo, W. Xiao, *Int. J. Hydrot. Energy* **2022**, *47*, 34583.
- [128] A. Agafonov, N. Pineda-Romero, M. Witman, V. Nassif, G. B. M. Vaughan, L. Lei, S. L. Ling, M. Dornheim, M. Allendorf, V. Stavila, C. Zlotea, *Acta Mater.* **2024**, *276*, 120086.
- [129] M. D. Witman, N. C. Bartelt, S. L. Ling, P. W. Guan, L. Way, M. D. Allendorf, V. Stavila, *J. Phys. Chem. Lett.* **2024**, *15*, 1500.
- [130] a) T. Sadhasivam, H. T. Kim, S. Jung, S. H. Roh, J. H. Park, H. Y. Jung, *Renew. Sust. Energ. Rev.* **2017**, *72*, 523; b) Q. Luo, J. D. Li, B. Li, B. Liu, H. Y. Shao, Q. Li, *J. Magnes. Alloy.* **2019**, *7*, 58; c) X. Lu, L. T. Zhang, H. J. Yu, Z. Y. Lu, J. H. He, J. G. Zheng, F. Y. Wu, L. X. Chen, *Chem. Eng. J.* **2021**, *422*, 130101.
- [131] a) L. Ren, Y. H. Li, N. Zhang, Z. Li, X. Lin, W. Zhu, C. Lu, W. J. Ding, J. X. Zou, *Nano-Micro Lett.* **2023**, *15*, 93; b) V. A. Yartys, M. V. Lototskyy, E. Akiba, R. Albert, V. E. Antonov, J. R. Ares, M. Baricco, N. Bourgeois, C. E. Buckley, J. M. B. von Colbe, J. C. Crivello, F. Cuevas, R. V. Denys, M. Dornheim, M. Felderhoff, D. M. Grant, B. C. Hauback, T. D. Humphries, I. Jacob, T. R. Jensen, P. E. de Jongh, J. M. Joubert, M. A. Kuzovnikov, M. Latroche, M. Paskevicius, L. Pasquini, L. Popilevsky, V. M. Skripnyuk, E. Rabkin, M. V. Sofianos, et al., *Int. J. Hydrot. Energy* **2019**, *44*, 7809; c) H. Y. Wan, X. Yang, S. M. Zhou, L. Ran, Y. F. Lu, Y. A. Chen, J. F. Wang, F. S. Pan, *J. Mater. Sci. Technol.* **2023**, *149*, 88; d) Q. Li, Y. F. Lu, Q. Luo, X. H. Yang, Y. Yang, J. Tan, Z. H. Dong, J. Dang, J. B. Li, Y. Chen, B. Jiang, S. H. Sun, F. S. Pan, *J. Magnes. Alloy.* **2021**, *9*, 1922; e) K. Wang, X. Zhang, Y. F. Liu, Z. H. Ren, X. L. Zhang, J. J. Hu, M. X. Gao, H. G. Pan, *Chem. Eng. J.* **2021**, *406*, 126831; f) S. J. Huang, M. P. Mose, S. Kannaiyan, *J. Compos. Sci.* **2021**, *5*, 145.
- [132] T. Malinova, Z. X. Guo, *Mat. Sci. Eng. A-Struct.* **2004**, *365*, 219.
- [133] Q. Jiang, Z. P. Jia, S. Lu, P. Song, Z. K. Gao, Z. G. Wang, T. R. Peng, X. Bai, H. Cui, W. Z. Tian, R. Feng, Q. Kang, Z. Y. Liang, H. K. Yuan, *Ceram. Int.* **2024**, *50*, 23058.
- [134] Q. F. Tian, Y. Zhang, Y. X. Wu, Z. C. Tan, *Int. J. Hydrot. Energy* **2009**, *34*, 1931.
- [135] N. Wang, S. P. Huang, *Phys. Rev. B* **2020**, *102*, 094111.
- [136] N. Wang, Beijing University of Chemical Technology, Beijing **2020**.
- [137] A. Pedrielli, P. E. Trevisanutto, L. Monacelli, G. Garberooglio, N. M. Pugno, S. Taioli, *Nanoscale* **2022**, *14*, 5589.
- [138] J. Lee, D. Sung, Y. K. Chung, S. Bin Song, J. Huh, *Nanoscale Adv* **2022**, *4*, 2332.
- [139] S. Y. Dong, Y. Y. Wang, J. Y. Li, Y. Y. Li, L. Wang, J. L. Zhang, *Int. J. Hydrot. Energy* **2023**, *48*, 38412.
- [140] K. Batalovic, J. Radakovic, B. P. Mamula, B. Kuzmanovic, M. M. Ilic, *Adv. Theory Simul.* **2022**, *5*, 2200293.
- [141] M. Witman, M. Allendorf, V. Stavila, **2022**.
- [142] a) H. Luo, Y. Yang, L. Lu, G. Li, X. Wang, X. Huang, X. Tao, C. Huang, Z. Lan, W. Zhou, J. Guo, H. Liu, *Appl. Surf. Sci.* **2023**, *610*, 155581; b) X. Zhang, W. X. Zhang, L. C. Zhang, Z. G. Huang, J. J. Hu, M. X. Gao, H. G. Pan, Y. F. Liu, *Chem. Eng. J.* **2022**, *428*, 132566; c) M. H. Wu, M. X. Gao, S. Q. Qu, Y. F. Liu, W. P. Sun, C. Liang, X. Zhang, Z. L. Li, Y. X. Yang, H. G. Pan, *J. Energy Storage* **2023**, *72*, 108679; d) X. C. Wang, X. Z. Xiao, J. G. Zheng, Z. M. Hang, W. P. Lin, Z. D. Yao, M. Zhang, L. X. Chen, *Int. J. Hydrot. Energy* **2021**, *46*, 23737.
- [143] a) Z. Ding, H. Li, L. Shaw, *Chem. Eng. J.* **2020**, *385*, 123856; b) X. Wang, Y. Jia, X. Xiao, P. Zhou, J. Bi, J. Qi, L. Lv, F. Xu, L. Sun, L. Chen, *J. Mater. Sci. Technol.* **2023**, *146*, 121; c) M. C. Wang, L. Z. Ouyang, M. Q. Zeng, J. W. Liu, C. H. Peng, H. Y. Shao, M. Zhu, *Int. J. Hydrot. Energy* **2019**, *44*, 7392.
- [144] R. Cheng, X. Y. Xue, C. Z. Wang, *J. Phys. Chem. C* **2024**, *128*, 7742.
- [145] a) D. Zhu, Y. Ait-Amirat, A. N'Diaye, A. Djerdir, *J. Energy Storage* **2021**, *42*, 102950; b) Y. Wang, H. Dai, H. Cao, S. Zhou, M. Gao, F. Sun, J. Liu, K. Han, J. Jiang, *Energy* **2024**, *305*, 132409; c) Y. Wang, H. Zhang, S. He, W. Wang, M. Gao, E. Koltsova, V. Vasilenko, *Energ. Convers. Mange.* **2023**, *291*, 117339; d) P. K. Kanti, A. P. Shrivastav, P. Sharma, M. P. Maiya, *Int. J. Hydrot. Energy* **2024**, *52*, 470.
- [146] J. Gao, X. M. Guo, Y. F. Wu, W. Xiao, L. Hao, *Int. J. Hydrot. Energy* **2024**, *58*, 657.
- [147] C. C. Peng, X. Y. Liu, R. Long, Z. C. Liu, W. Liu, *Chem. Eng. Res. Des.* **2024**, *207*, 100.
- [148] C. S. Wang, J. Brinkerhoff, *Int. J. Hydrot. Energy* **2021**, *46*, 24256.
- [149] S. H. Suárez, D. Chabane, A. N'Diaye, Y. Ait-Amirat, A. Djerdir, *Int. J. Hydrot. Energy* **2024**, *50*, 178.
- [150] P. Zhao, X. G. Zeng, H. Q. Kou, H. Y. Chen, *Int. J. Hydrot. Energy* **2024**, *54*, 1251.
- [151] N. Mounet, M. Gibertini, P. Schwaller, D. Campi, A. Merkys, A. Marrazzo, T. Sohier, I. E. Castelli, A. Cepellotti, G. Pizzi, N. Marzari, *Nat. Nanotechnol.* **2018**, *13*, 246.
- [152] a) H. J. Lin, Y. S. Lu, L. T. Zhang, H. Z. Liu, K. Edalati, A. Révész, *Rare Met.* **2022**, *41*, 1797; b) W. F. Qiao, D. M. Yin, S. L. Zhao, N. Ding, L. Liang, C. L. Wang, L. M. Wang, M. He, Y. Cheng, *Chem. Eng. J.* **2023**, *465*, 142837; c) P. P. Zhou, Z. M. Cao, X. Z. Xiao, R. H. Li, Z. Q. Liang, H. K. Zhang, L. J. Zhan, Z. N. Li, L. J. Jiang, L. X. Chen, *ACS Appl. Energy Mater.* **2022**, *5*, 3783; d) R. R. Shahi, A. K. Gupta, P. Kumari, *Int. J. Hydrot. Energy* **2023**, *48*, 21412.
- [153] a) J. P. Bi, P. P. Zhou, X. Z. Xiao, Y. J. Zhang, H. Q. Kou, T. Tang, L. X. Chen, *Chem. Eng. J.* **2024**, *481*, 148517; b) Y. Madaria, E. A. Kumar, S. S. Murthy, *Appl. Therm. Eng.* **2016**, *98*, 976; c) C. H. Li, X. L. Wang, J. M. Wu, C. Y. Wang, *J. Power Sources* **1998**, *70*, 106; d) Y. Y. Zhu, L. Z. Ouyang, H. Zhong, J. W. Liu, H. Wang, H. Y. Shao, Z. G. Huang, M. Zhu, *Angew. Chem., Int. Ed.* **2020**, *59*, 8623.
- [154] a) Y. Reich, N. Travitzky, *Mater. Des.* **1995**, *16*, 251; b) Y. Zhang, C. Ling, *NPJ Comput. Mater.* **2019**, *3*, 71; c) H. Shimakawa, A. Kumada, M. Sato, *NPJ Comput. Mater.* **2024**, *10*.
- [155] a) R. Narikawa, Y. Fukatsu, Z. L. Wang, T. Ogawa, Y. Adachi, Y. Tanaka, S. Ishikawa, *Adv. Theory Simul.* **2022**, *5*, 2100470; b) D. C. Li, I. H. Wen, *Neurocomputing* **2014**, *143*, 222; c) A. Berrones, E. Jiménez, M. A. Alcorta-García, F. J. Almaguer, B. Peña, *Neurocomputing* **2016**, *175*, 555.
- [156] H. Ohno, *Appl. Soft Comput.* **2020**, *86*, 105932.
- [157] Y. X. Han, Y. H. Liu, Q. C. Chen, *Comp. Mater. Sci.* **2023**, *226*, 112250.
- [158] C. Allen, S. Aryal, T. Do, R. Gautum, M. M. Hasan, B. K. Jasthi, E. Gnimpieba, V. Gadhamshetty, *Front. Microbiol.* **2022**, *13*, 1059123.
- [159] a) T. Lookman, V. Balachandran Prasanna, D. Xue, R. Yuan, *NPJ Comput. Mater.* **2020**, *4*, 25; b) D. Z. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Q. Xue, T. Lookman, *Nat. Commun.* **2016**, *7*, 11241; c) D. Z. Xue, D. Q. Xue, R. H. Yuan, Y. M. Zhou, P. V. Balachandran, X. D. Ding, J. Sun, T. Lookman, *Acta Mater.* **2017**, *125*, 532.
- [160] S. Z. Chen, H. Cao, Q. B. Ouyang, X. Wu, Q. Qian, *Mater. Des.* **2022**, *223*.
- [161] a) F. Z. Zhuang, Z. Y. Qi, K. Y. Duan, D. B. Xi, Y. C. Zhu, H. S. Zhu, H. Xiong, Q. He, *Proc. IEEE* **2021**, *109*, 43; b) V. Cheplygina, M. de Brujinne, J. P. W. Pluim, *Med. Image Anal.* **2019**, *54*, 280; c) M. S. Long,

- J. M. Wang, G. G. Ding, S. J. Pan, P. S. Yu, *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 1076.
- [162] C. Zhu, J. Y. Ni, Z. Yang, Y. Sheng, J. Yang, W. Zhang, *Comput. Theor. Chem.* **2022**, 1217.
- [163] a) S. Velliangiri, S. Alagumuthukrishnan, S. I. Thankumar joseph, *Procedia Comput. Sci.* **2019**, *165*, 104; b) G. Huang, S. J. Song, J. N. D. Gupta, C. Wu, *IEEE Trans. Cybern.* **2014**, *44*, 2405; c) R. Sheikhpour, M. A. Saram, S. Gharaghani, M. A. Z. Chahooki, *Pattern Recogn.* **2017**, *64*, 141.
- [164] D. L. Chen, D. P. Sun, J. Fu, S. X. Liu, *IEEE Access* **2021**, *9*, 30858.
- [165] a) Y. Luo, S. F. Chen, G. Valdes, *Med. Phys.* **2020**, *47*, E178; b) T. Mori, N. Uchihira, *Empir. Softw. Eng.* **2019**, *24*, 779.
- [166] a) J. H. Huang, W. Z. Sun, L. Huang, *Neurocomputing* **2020**, *378*, 260; b) T. T. Le, W. X. Fu, J. H. Moore, *Bioinformatics* **2020**, *36*, 250; c) T. Torneide, A. Tornede, J. Hanselle, F. Mohr, M. Wever, E. Hüllermeier, *J. Artif. Intell. Res.* **2023**, *77*, 427.
- [167] a) A. Singh, K. Gaurav, *Sci. Rep.* **2023**, *13*, 2251; b) J. Waring, C. Lindvall, R. Umeton, *Artif. Intell. Med.* **2020**, *104*, 101822; c) C. Thongprayoon, P. Pattharanitima, A. G. Kattah, M. C. A. Mao, M. T. Keddis, J. J. Dillon, W. Kaewput, S. Tangpanithandee, P. Krisanapan, F. Qureshi, W. Cheungpasitporn, *J. Clin. Med.* **2022**, *11*, 6264; d) T. Lazebnik, T. Fleischer, A. Yaniv-Rosenfeld, *Sustainability* **2023**, *15*, 11232; e) F. Conrad, M. Mälzer, M. Schwarzenberger, H. Wiemer, S. Ihlenfeldt, *Sci. Rep.* **2022**, *12*, 19350.
- [168] B. C. Sun, W. J. Cui, G. Y. Liu, B. Zhou, W. J. Zhao, *Case Stud. Constr. Mater.* **2023**, *19*, e02405.
- [169] a) J. D. O'Shea, K. A. Crockett, W. Khan, Z. Bandar, *International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, BRAZIL, **2018**; b) S. C. Jia, P. W. Lin, Z. Y. Li, J. W. Zhang, S. X. Liu, *J. Visualization* **2020**, *23*, 141; c) X. X. Niu, C. Y. Suen, *Pattern Recogn.* **2012**, *45*, 1318; d) A. A. Peres, S. M. Vieira, J. R. C. Pinto, *International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, BRAZIL, **2018**.
- [170] S. Ahlawat, A. Choudhary, *International Conference on Computational Intelligence and Data Science (ICCIDS)*, NorthCap Univ, Gurugram, India, Sep 06–07, **2019**.
- [171] a) D. Mundici, presented at *International School on Neural Nets E R Caianiello 5th Course: From Synapses to Rules – Discovering Symbolic Rules From Neural Processed Data*, Erice, Italy, Feb 25–Mar 07, **2002**; b) J. Denker, D. Schwartz, B. Wittner, S. Solla, R. Howard, L. Jackel, J. Hopfield, *Complex Sys* **1987**, *1*, 877; c) G. Bologna, *Int. J. Neural Syst.* **2001**, *11*, 247.
- [172] A. E. A. Allen, A. Tkatchenko, *Sci. Adv.* **2022**, *8*, eabm7185.
- [173] a) G. Visani, E. Bagli, F. Chesani, A. Poluzzi, D. Capuzzo, *J. Oper. Res. Soc.* **2022**, *73*, 91; b) D. Mane, A. Magar, O. Khode, S. Koli, K. Bhat, P. Korade, *J. Electr. Sys.* **2024**, *20*, 1252; c) G. D. Torres, M. F. Roig-Maimó, M. M. Mascaró-Oliver, E. Amengual-Alcover, R. Mas-Sansó, *Sensors* **2023**, *23*, 131.



**Panpan Zhou** received his Ph.D. degree in materials science from Zhejiang University, China, in 2024. He is currently a lecturer in the College of Materials Science and Engineering at Hohai University, China. His research interests include medium- and high-entropy hydrogen storage materials customization, poisoning mechanisms and anti-poisoning modifications, hydrogen evolution catalysts synthesis, first principles calculation, and artificial intelligence.



**Qianwen Zhou** obtained her B.S. degree from the College of Materials Science and Engineering at Hohai University, China, in 2023. Since September 2023, she has been a doctoral candidate in the School of Materials Science and Engineering, Zhejiang University. Her research interests include poisoning mechanisms and anti-poisoning modifications of hydrogen storage materials, and first principles calculation.



**Xuezhang Xiao** is an associate professor at the School of Materials Science and Engineering at Zhejiang University, China. He is a member of China Renewable Energy Society. His main research interests focus on the fundamental exploration and practical applications of renewable energy materials, including novel metal hydrides and complex hydrides for hydrogen storage, nano/amorphous hydrogen storage materials and devices for fuel cell and Ni/MH batteries, and high throughput design.



**Xiulin Fan** is a professor at the School of Materials Science and Engineering, Zhejiang University. His research interests include design of metal hydride and complex hydride materials, interfacial engineering of energy storage devices such as lithium-ion batteries, sodium-ion batteries, potassium-ion batteries, etc., and theoretical design of related electrolytes, high-throughput mining.



**Yongjin Zou** is a full professor at the School of Materials Science and Engineering, Guilin University of Electronic Technology. His research interests include hydrogen storage materials, gas-sensitive materials, phase change materials, electrode materials, as well as theoretical calculations and database construction related to material design.



**Lixian Sun** is a full professor at the School of Materials Science and Engineering, Guilin University of Electronic Technology. He has been elected as a Foreign Member of Russian Academy of Sciences and a Foreign Member of the European Academy of Sciences. His research interests encompass materials for new energy (hydrogen storage, hydrogen production, and phase change energy storage), (bio)fuel cells, nickel-metal hydride/lithium-ion batteries, novel catalysts, carbon dioxide capture, as well as theoretical calculations and database construction related to material design.



**Jinghua Jiang** is a full professor at the College of Materials Science and Engineering, Hohai University. She is a lifetime member of the Chinese Materials Research Society. Her research interests include preparation of nano/ultrafine/heterostructural alloys and their structure-efficiency relationships, structure-function integrated light alloys and their surface technology, and hydrogen storage materials with their application design.



**Dan Song** is a full professor at the College of Materials Science and Engineering, Hohai University. His main research interests include high-performance ultrafine/nano-metallic materials, corrosion and protection of materials, surface engineering of materials, corrosion and control of steel reinforcement in concrete, and hydrogen storage materials with their application design.



**Lixin Chen** is a full professor at the School of Materials Science and Engineering, Zhejiang University. He also serves as the deputy director of the Hydrogen Energy Professional Committee of China Renewable Energy Society, a member of the Chinese National Standardization Technical Committee for Hydrogen Energy. His research interests focus on the fundamental research and applications of energy storage materials, including hydrogen storage materials, metal hydride technologies for hydrogen storage/compression/purification, hydrogen isotope separation and storage systems, novel battery materials, etc.