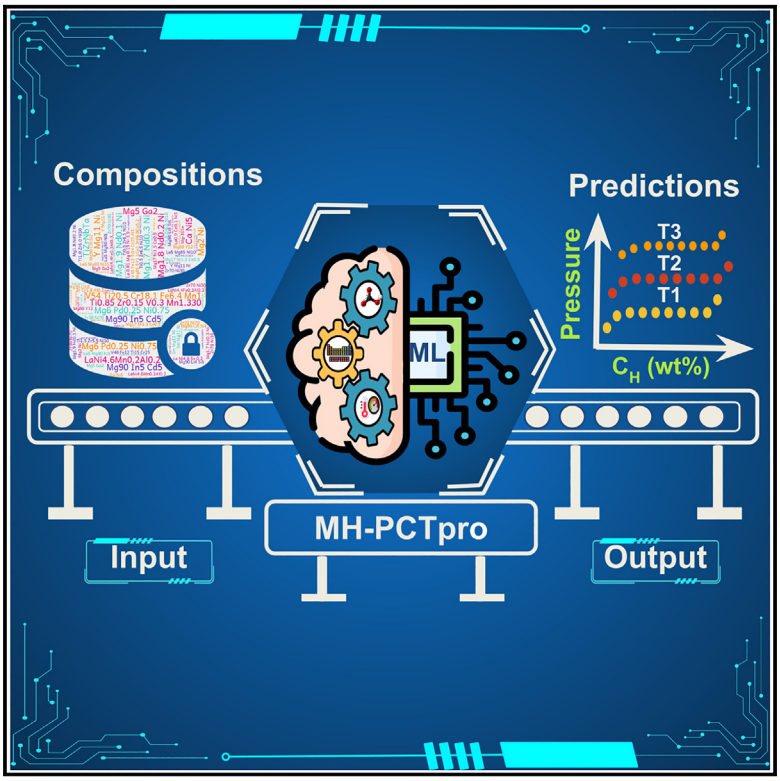
iScience

Article

MH-PCTpro: A machine learning model for rapid prediction of pressure-composition-temperature (PCT) isotherms

Graphical abstract



Authors

Ashwini Verma, Kavita Joshi

Correspondence

k.joshi@ncl.res.in

In brief

Machine learning; Computational materials science; Energy materials

Highlights

- MH-PCTpro, an ML model predicts PCT isotherms for metal compositions
- Trained on 14,000+ data points from 237 PCT isotherms across 138 compositions
- Features include elemental and hydriding properties, plus experimental parameters
- Provides thermodynamic insights via temperature-dependent PCT isotherm prediction

Verma & Joshi, 2025, iScience *28*, 112251
April 18, 2025
https://doi.org/10.1016/j.isci.2025.112251

CellPress

iScience

CellPress
OPEN ACCESS

Article

MH-PCTpro: A machine learning model for rapid prediction of pressure-composition-temperature (PCT) isotherms

Ashwini Verma[1,2] and Kavita Joshi[1,2,3,*]
[1]Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pashan, Pune 411008, India
[2]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India
[3]Lead contact
*Correspondence: k.joshi@ncl.res.in
https://doi.org/10.1016/j.isci.2025.112251

SUMMARY

We present a machine-learning powered Metal Hydride's Pressure-Composition-Temperature isotherm Predictor (MH-PCTpro) for metal compositions. To train the MH-PCTpro, an experimental database of PCT isotherms is built from published literature. The database comprises over 14,000 data points extracted from 237 PCT isotherms representing 138 distinct compositions. The dataset encompasses more than 25 elements and spans a broad spectrum of absorption temperatures (263–653 K) and hydrogen pressures (0.001–40 MPa). The model is validated on a wide range of alloy families and its predictions are consistent with experimental results. The model also captures temperature-dependent variations in plateau pressure, enabling determination of enthalpy and entropy of hydride formation through Van't Hoff plots. Hence, MH-PCTpro can be used as an ML tool for guiding PCT experiments, offering PCT isotherm predictions and valuable thermodynamic insights into materials suitable for solid-state hydrogen storage.

INTRODUCTION

Solid-state hydrogen storage in metal alloys holds the potential to outperform conventional (liquid/compressed) methods on accounts of volumetric capacity, safety, and durability.[1–3] One major advantage of hydrogen storage in metal hydrides is their energy-efficient way of storing hydrogen at relatively low pressures, and moderate temperatures eliminating the need for liquefaction or compression. The metal-hydrogen bond offers the advantage of a high volumetric hydrogen density under moderate pressures.[4,5] Despite significant performance improvements in metal hydrides over the past two decades, their practical application is hindered by the unattainable thermodynamic and kinetic properties at ambient conditions for economic viability.[6,7] The storage properties of metal alloys are strongly related to the hydrogen pressure, temperature, and the alloy compositions. This presents a vast chemical space that needs to be explored to discover suitable solid-state hydrogen storage materials.

The PCT analysis is employed to assess the suitability of an alloy for hydrogen storage. The PCT isotherm, also known as a pressure-composition-isotherm (PCI) curve, illustrates how the hydrogen content in a material changes with pressure and temperature. The PCT isotherms of an alloy are determined by conducting thermogravimetric and volumetric measurements with Sievert-type apparatus.[8] They provide crucial information, like maximum reversible storage capacity, hydride phase transition, and the equilibrium plateau pressure. The length of plateau determines how much hydrogen can be stored reversibly. The relationship between the plateau pressure (P) and temperature (T) follows the Van't Hoff equation, enabling determination of enthalpy and entropy changes associated with the de/hydriding reactions. Hence, PCT characterization provides insights into essential hydrogen storage properties of materials.[9]

Unfortunately, PCT analysis is resource-intensive and time-consuming as it involves a series of measurements to represent the relationship between hydrogen pressure, concentration, and temperature of a sample at equilibrium, and hence limiting the number of compositions that can be investigated. Consequently, there is a need for alternative methods to expedite the determination of PCT isotherms for metal alloys, enabling the rapid exploration of materials suitable for solid-state hydrogen storage. Recently, Zepon et al. proposed the first thermodynamic model for calculating PCT diagrams of single-phase BCC multi-component alloys.[10] The model is based on determining Gibbs free energy of the metal-hydrogen system using (1) the entropy of mixing of hydrogen in the alloy phases (BCC phase and hydride phase (α and β)), (2) the enthalpy of hydrogen mixing, and (3) the chemical potentials of hydrogen and metal. The enthalpy terms for constituent elements of alloy were parameterized by combining experimental data and total energy calculations through density functional theory (DFT). They validated the model by comparing the calculated PCT isotherms of various BCC multicomponent alloys with experimentally reported PCT isotherms. Building on this, Ponsoni et al. employed the same thermodynamic model to predict PCT isotherm of multicomponent alloys having C14 Laves phase.[11] For both

Table 1. The average MAE and R2 score for series of evaluations, including an 80:20 data split of (1) 14000 points, (2) 237 PCT isotherms, and (3) 138 compositions over 100 iterations

DataSet	MAE_Train	MAE_Validation	R2 Score
80:20 split of 14000 data points	0.12 ± 0.002	0.17 ± 0.002	0.96
80:20 split of 237 PCT isotherms	0.11 ± 0.006	0.55 ± 0.08	0.64
80:20 split of 138 compositions	0.11 ± 0.009	0.64 ± 0.13	0.55

BCC and C14 Laves phase multicomponent alloys the calculated plateau pressures were found to be in close proximity to experimental values. However, the model requires further improvement to accurately predict plateau lengths. Additionally, the model's dependence on DFT calculations and experimental data related to hydrogen-alloy phases poses constraints on its generalizability and predictability as well.

Machine learning (ML) models offer a host of advantages over thermodynamic models, particularly when dealing with complex problems. One of the most significant advantages is their ability to learn valuable insights from data without requiring *a priori* knowledge of physical laws or equations.[12] Today, ML is playing a crucial role in expediting the discovery, design, and development of materials.[13,14] In the field of solid-sate hydrogen storage, various ML models have been employed to predict properties like hydrogen weight capacity,[15] enthalpy of hydride formation,[15–19] plateau pressure,[19,20] etc. These models demonstrated the advantages of ML to gain insights and predicting hydrogen storage properties using simple elemental properties. Some of these initial models, trained on a vast, auto-generated feature set, offer insights into factors affecting hydrogen storage in metal alloys. While these interpretable models enhance understanding, their complexity can affect prediction accuracy.[21] Subsequently, models are refined by improving datasets and adding relevant features resulting in significant improvement in learning and prediction accuracy.[15,22] Our recent models have included temperature as one of the parameters bringing predictions closer to reality. Some of these models are also employed to predict hydrogen storage properties across a wide range of chemical compositions in metal alloys, facilitating the identification of potential candidates for efficient hydrogen storage.[15] It is important to note that despite their utility, all models have limitations. The scarcity of data is one of the significant limitations. Some models reveal that if trained on a subset of data, predictions on the same subset are impressive.[16] However, transferability of these models to new material classes and compositions is limited.

Recently, Kim et al. reported a ML model for predicting PCT curves of AB_2-type hydrogen storage alloys.[23] They trained three different algorithms: K-Nearest Neighbors (KNN), Random Forest (RF), and Deep Neural Networks (DNN) using PCT data from 33 different AB_2 alloys. To improve on prediction accuracy, they employed fitting functions to generate additional data points from the experimental PCT data. Additionally, they have employed the Van't Hoff equation to generate unmeasured temperature data. The model's predictions of PCT curves were evaluated at fixed as well as arbitrary temperatures, and the results aligned with experimentally reported values. The features used for training are the compositional fraction of elements present in the alloys and absorption pressures. Just taking compositional fractions might impose limitation when working with larger pool of compositions as the fractions do not provide any information about the properties of the materials. Also, the model cannot make PCT predictions for various classes of alloys across a wide temperature range, primarily due to its training on limited data. Therefore, the need of an efficient and generalized model remains an open challenge.

In light of this, we present an ML model called MH-PCTpro to predict the PCT isotherms for various metal hydrides. The model learns from existing experimental data of diverse metal compositions, identifying relationships between material properties and their PCT isotherms. The designed feature set includes experimental parameters like absorption temperature and pressure along with the periodic table and hydriding properties of elements. MH-PCTpro, trained on 237 PCT isotherms, achieved an average MAE of 0.17 ± 0.002 wt % and an R2 score of 0.96. We evaluated the model's accuracy using unseen datasets and found it effective for alloys containing various elements, including light metals, transition metals, and lanthanides. Additionally, we tested the model's performance using smaller datasets and evaluated solutions for improving its accuracy with limited data. These results highlight that the ML models based on even sparse but high-quality data along with chemically relevant descriptors effectively predict complex material characteristics like PCT isotherms to accelerate the discovery of high-performance materials.

RESULTS AND DISCUSSION

MH-PCTpro predicts PCT isotherms for various compositions as a function of hydrogen pressure and absorption temperature. To scrutinize the model's robustness and accuracy in predicting the PCT isotherm, a series of evaluations are conducted with different subsets of data and features. First, random 80:20 train-validation split is performed on the approximately 14,000 data points to train MH-PCTpro. Within this data distribution, MH-PCTpro yielded an average MAE 0.17 ± 0.002 wt % of validation set and an impressive R2 score of 0.96 as shown in Table 1. We conducted K-fold cross-validation to assess the model's performance on unseen data as shown in Figure S1. The accuracy achieved through K-fold cross-validation is comparable to that obtained with an 80:20 train-validation split. Additionally, the comparison between the experimental values and those predicted by the ML model is shown in Figure S2. MH-PCTpro, trained and tested on a diverse dataset of PCT isotherms from various alloy classes, achieves comparable accuracy to model reported by Kim et al.[23] (R2 Score = 0.98), which was limited to PCT isotherms of a single alloy class.

Next, we perform an 80:20 train-validation split on 237 PCT isotherms, resulting in approximately 190 PCT isotherms in the training set and the remaining 47 PCT isotherms in the validation set. Similarly, the same split is applied to 138 unique compositions, yielding approximately 100 compositions in the training set and the remaining 38 compositions in the validation set for

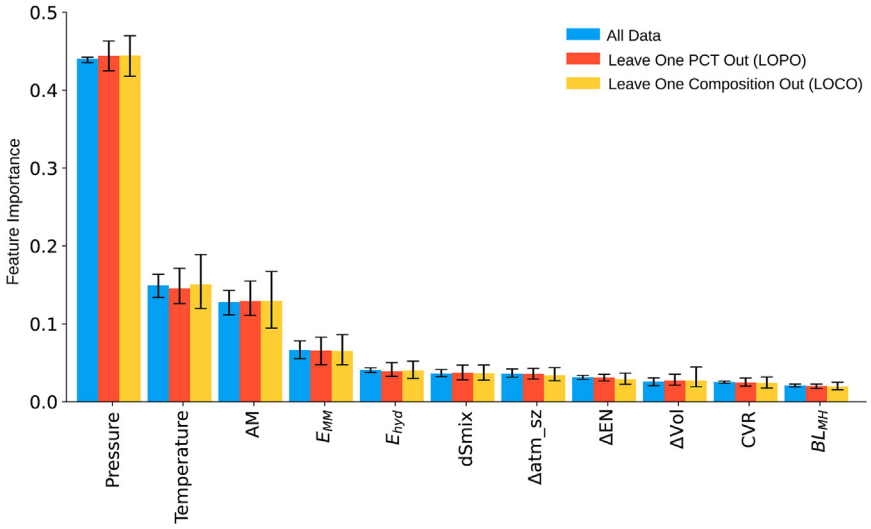


Figure 1. Feature importance identified by the ETR model for PCT isotherm prediction over 100 iterations

Hydrogen pressure, absorption temperature, atomic mass (AM), metal-metal dimer bond-energy (E_{MM}), entropy of mixing (dSmix), enthalpy of hydride formation (E_{hyd}), lattice distortion (Δ atm_sz) and electronegativity difference (Δ EN) are among the most essential features.The maximum deviation in feature importance measurement over 100 iterations is observed for leave-one-composition-out(LOCO). In the case of LOCO, the most important feature, hydrogen pressure, has a feature importance of 0.45 ± 0.03.

100 iterations. Our aim is to build a model which can predict PCT isotherm as a function of temperature for a given composition. We performed the latter two experiments by withholding 20% of the PCT isotherms/compositions from the training set to evaluate model performance in predicting complete PCT curve. The MAEs for the latter two cases are calculated by averaging the MAE values for each point on a PCT curve. For these two cases, the MAE increases and a corresponding decrease in the R2 score is noted (refer to Table 1). The observed drop in R2 score is attributed to the lack of precise overlap between every point of a predicted and actual PCT isotherms. It is important to note that error bars such as MAE and R2 score are typically used for point predictions. However, in the case of predicting the entire PCT isotherms, the model's predictability is not entirely evident from the magnitudes of the R2 score obtained from the average R2 score of all the predicted points on a PCT plot. We will discuss this in detail shortly.

Figure 1 illustrates the feature importance identified by the ETR model for three different input cases as discussed in Table 1 i.e., split over (1) all data points, (2) 237 PCT isotherms, and (3) 138 compositions. The order of important features across three distinct input data scenarios remains consistent, indicating the stability of the model's learning across varied inputs. Hydrogen pressure and absorption temperature are among the most essential features, emphasizing their substantial influence on the model's predictability. Furthermore, the importance of elemental features like atomic mass (AM), metal-metal dimer bond-energy (E_{MM}), entropy of mixing (dSmix), enthalpy of hydride formation (E_{hyd}), lattice distortion (Δ atm_sz), electronegativity difference (Δ EN), volume difference between the element's bulk phase and its hydride phase (Δ Vol), covalent radius (CVR), and metal-hydrogen dimer bond length (BL_{MH}) underscores the significance of fundamental properties of constituent elements of compositions in the model's learning.

We evaluated the impact of incorporating the new features, E_{hyd} and Δ Vol on MH-PCTpro predictability by individually predicting PCT isotherms for 138 compositions. Approximately

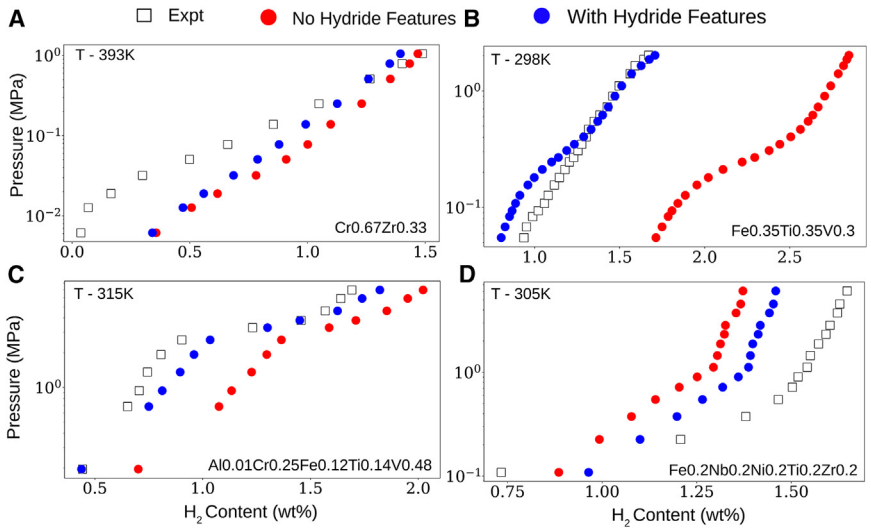


Figure 2. Comparison of predicted and experimental PCT isotherms

Black squares represents experimental PCT isotherms. Predicted isotherms with and without hydriding features are shown in blue and red circles, respectively. The experimental data is taken from ref. 24–26, and 53 for (A), (B), (C), and (D), respectively.

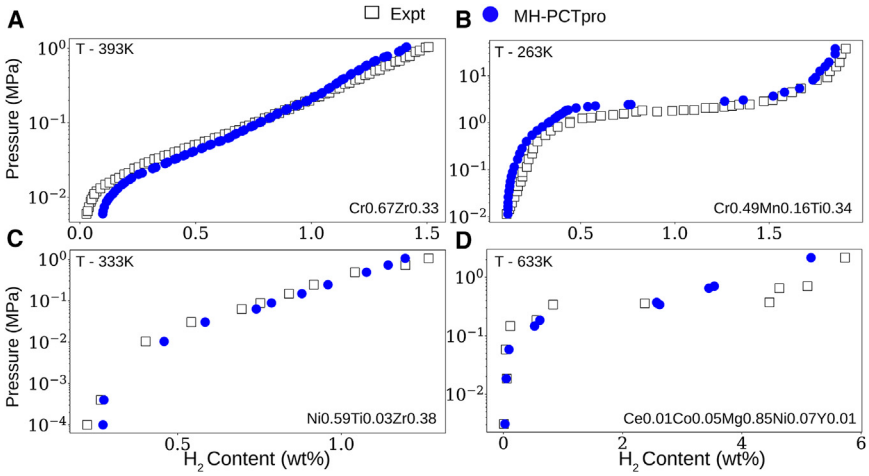


Figure 3. Accurately predicted PCT isotherms for Leave-One-PCT-Out analysis

The black squares represent the experimentally reported, and the blue circles represent predicted PCT isotherms. The experimental data is taken from ref. 24,27,28, and 29 for (A), (B), (C), and (D).

10% of the predicted PCT isotherms show significant improvement upon addition of hydriding properties as features, particularly for compositions comprising only transition metals (TM) with higher fractions. In Figure 2 the black squares indicate experimentally reported PCT isotherms, while the blue and red circles depict predicted PCT isotherms with and without the inclusion of hydriding properties as feature. In all four cases (A) Cr0.67Zr0.33, (B) Fe0.35Ti0.35V0.3, (C) Al0.01Cr0.25Fe0.12Ti0.14V0.48, and (D) Fe0.2Nb0.2Ni0.2Ti0.2Zr0.2, the compositions consist of hydriding elements like V, Ti, Zr, and Nb and non-hydriding elements Cr, Fe, Al, and Ni. Although V, Ti, Zr, Nb, Cr, Fe, Mo, Mn, and Ni all are transition metals, compositions consisting of these elements exhibit significantly different PCT isotherms. The variation in the PCT isotherms depends on the fractions of each constituent element present in the composition. For instance, the introduction of non-hydriding elements like Cr, Mn, and Fe (which have positive enthalpies of hydride formation), to stable hydrides of V, Ti, and Nb (which have negative enthalpies of hydride formation), lead to the destabilization of the resultant hydride. Hence, the features quantifying hydriding properties proved crucial in understanding the impact of hydriding and non-hydriding elements on the PCT isotherms of compositions consisting of only TM; leading to improved predictions as shown in Figure 2 (A), (B), (C), and (D) that align closely with experimental results. We have included hydriding properties into final training of MH-PCTpro and used for all the predictions shown hereafter. Examples of compositions wherein model predictions exhibit negligible or marginal changes following the inclusion of hydriding features are illustrated in Figure S3. These compositions predominantly comprise alkali metals, alkaline earth metals, non-metals, and lanthanides as primary constituent elements. Given the distinct properties of these elements across the periodic table, the model relies on elemental and compositional features to make predictions, resulting in minimal alterations upon the incorporation of hydriding features.

Understanding the MH-PCTpro

Understanding the predictions made by the ML model is crucial for assessing its applicability and limitations. To achieve this for MH-PCTpro, we have carried out two specific experiments. In the first experiment Leave-One-PCT-Out (LOPO), each PCT isotherm is predicted while retaining the remaining 236 PCT isotherms in the training set. Similarly, in the second experiment Leave-One-Composition-Out (LOCO), each composition's PCT isotherms are excluded from the training set at a time for prediction. Table S1 and Table S2 illustrate the error bars of the model's

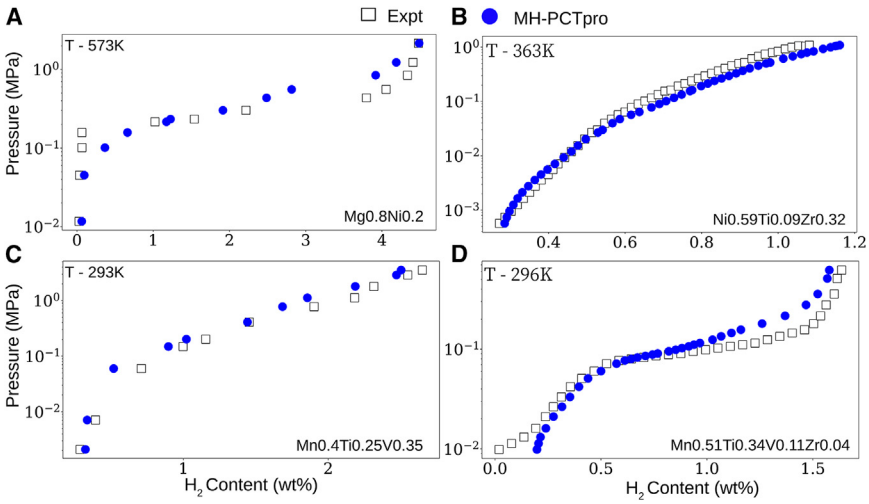


Figure 4. Accurately predicted PCT isotherms for Leave-One-Composition-Out analysis

The black squares represent the experimentally reported, and the blue circles represent predicted PCT isotherms. The experimental data is taken from ref. 24,30,31, and 32 for (A), (B), (C), and (D).

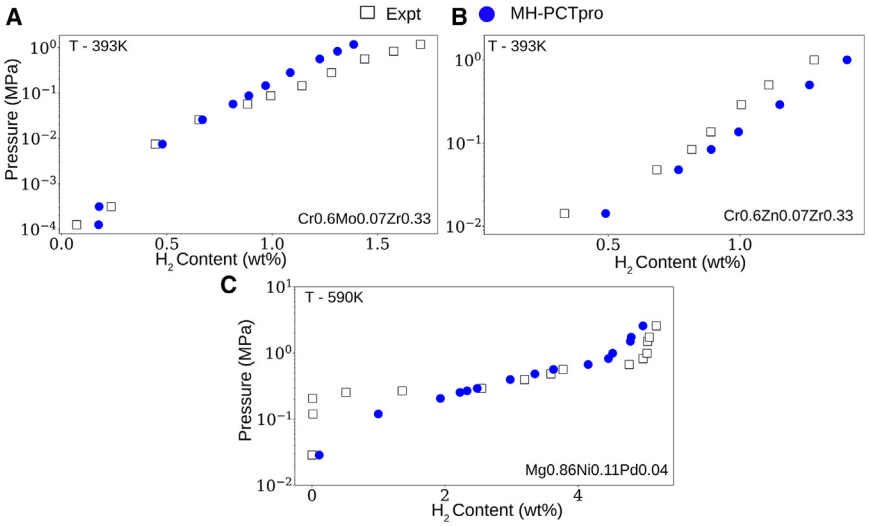


Figure 5. Comparison of experimentally reported (black squares) and predicted (blue circles) PCT isotherms for unseen transition metals

The experimental data is taken from ref. 24 for (A) and (B) and from ref. 33 for (C).

performance in isolated PCT isotherm predictions. For analysis, we categorized the predicted PCT isotherms into three classes based on the model's accuracy in predicting the overall shape of the PCT isotherms and the maximum weight capacity. (1) Successful prediction includes cases where the model accurately predicted both the shape of the curve and the maximum weight capacity. (2) Moderate predictions comprise instances where the model has either predicted the shape of the curve with reasonable accuracy or the maximum weight capacity well. (3) Poor prediction includes cases where the model has neither predicted the shape of the curve nor the maximum weight capacity well.

Leave-one-PCT-out

For LOPO, all the predicted isotherms fall under successful prediction criteria i.e., model has accurately predicted both the shape of the curve and the maximum weight capacity. Figure 3 (A) Cr0.67Zr0.33, (B) Cr0.49Mn0.16Ti0.34, (C) Ni0.59Ti0.03Zr0.38, and (D) Ce0.01Co0.05Mg0.85Ni0.07Y0.01 showcases examples from best-predicted isotherms. These predictions demonstrate the model's ability to extrapolate effectively at different temperatures, provided some relevant isotherms should be present in the training data. However, in practice the ML model is required to predict PCT isotherms for entirely new compositions not included in the training dataset. To assess the MH-PCTpro in this regard, we performed Leave-One-Composition-Out (LOCO), enabling the evaluation of the model's predictability for unseen compositions.

Leave-one-composition-out

When the model is retrained by leaving one composition out and used for predicting the PCT isotherms of left out composition, 80% of the compositions met the norm for "successful prediction", 14% fall under "moderate prediction", and only 6% predictions could not reproduce the shape of PCT or the maximum H2wt %, thus falling under the category of "poor predictions". It is important to note that in case of LOCO, no prior information is available in the dataset about the specific composition. Yet model has successfully predicted 80% of the PCT isotherms at different temperatures. Figure 4 (A) Mg0.8Ni0.2, (B) Ni0.59Ti0.09Zr0.32, (C) Mn0.4Ti0.25V0.35, and (D) Mn0.51Ti0.34V0.11Zr0.04 represent the accurately predicted PCT isotherms. The prediction for these four different compositions is consistent with the experimental results in terms of both the shape of the PCT isotherm and the maximum weight capacity. All the accurately predicted compositions comprise of alkali metals, alkaline earth metals, TM, and some lanthanides as constituent elements. The model's strong predictability for PCT isotherms of new compositions containing these elements can be attributed to the training data. The training data comprised a substantial proportion of TM and Mg-based compositions, along with a fair amount of lanthanide-based compositions. Interestingly, the model's capabilities extend beyond the elements included in the training data. Compositions having TMs such as Mo, Pd, and Zn are new to the model, yet the predicted PCT isotherms closely align with experimental results, as depicted in the Figure 5A Cr0.6Mo0.07Zr0.33, (B) Cr0.6Zn0.07Zr0.33, and (C) Mg0.86Ni0.11Pd0.04. Similarly, the model has also predicted moderately well for compositions having unseen light-weight hydriding element Li with Mg as depicted in Figure 6 (A) Li0.1Mg0.9 and (B) Li0.04Mg0.9Si0.06.

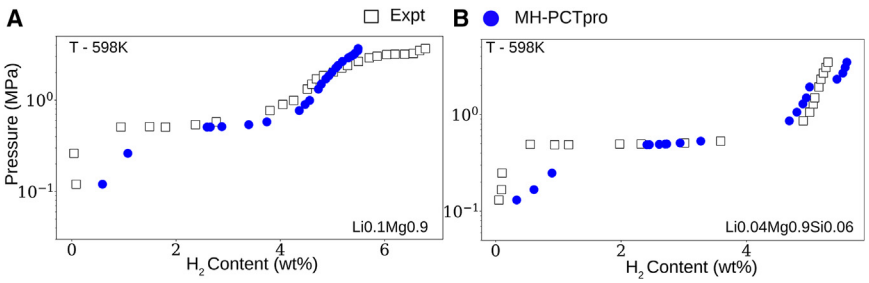


Figure 6. Comparison of experimentally reported (black squares) and predicted (blue squares) PCT isotherms for unseen elements Li and Si

The experimental data is taken from ref.34 for (A) and (B).

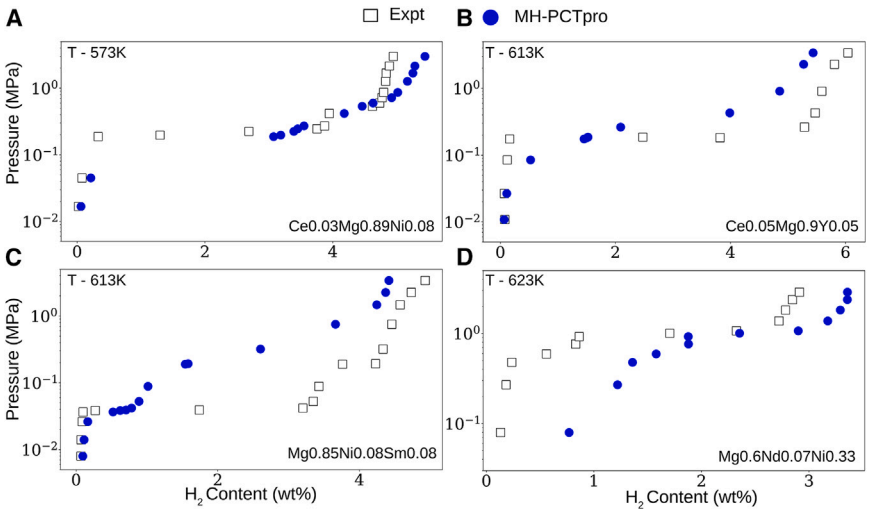


Figure 7. Comparison of experimental and predicted PCT isotherms for compositions containing unseen lanthanides like Ce, Sm, and Nd

The experimental data is taken from ref.35–37 and 38 for (A), (B), (C), and (D), respectively.

Figure 7 shows the predicted PCT isotherms of compositions containing lanthanides (Ce, Sm, and Nd) as dopants. On a closer look at predicted PCT isotherms of compositions shown in Figure 7 (A) Ce0.03Mg0.89Ni0.08, (C) Mg0.85Ni0.08Sm0.08 and (D) Mg0.6Nd0.07Ni0.33, it becomes apparent that, despite sharing the same base family (Mg-Ni), the experimentally observed PCT isotherms and plateau pressure positions vary significantly depending on the dopant. It is worth highlighting that MH-PCTpro effectively predicts the changes in the characteristics of PCT isotherms induced by the different dopants. The training set comprises 37% of PCT isotherms of lanthanides (La and Ce), measured at wide range of temperatures, enables the model to accurately predict for the new compositions consisting these elements (La and Ce) as shown in Figure 7 (A) and (B). However, for compositions consisting heavier lanthanides the predictions fall under "moderate predictions" as shown in Figure 7 (C) and (D).

Also for some less-represented binary families in training set like Cr-Zr, Mg-V and high entropy alloys the predictions falls under "moderate predictions" as shown in Figure S4. Only a small fraction (6%) of predictions fall within the poor performance category. Figure 8 (A) In0.02Mg0.98, (B) Ga0.29Mg0.71, (C) La0.24Mg0.73Ni0.02, and (D) Al0.07-La0.17Ni0.77 show experimental and predicted PCT isotherms for four different compositions. Model has failed in accurately predicting PCT isotherms of these compositions. In the worst predicted PCT isotherms, the compositions either possess less-represented class of constituent elements like non-metals (In and Ga) as shown in Figures 8A and 8B or fall within under-represented temperature range (400-500K) in the training data as shown in Figures 8C and 8D.

Temperature dependent PCT prediction

The influence of temperature on hydrogen storage properties is assessed by comparing PCT isotherms at various temperatures. MH-PCTpro is trained on diverse compositions with PCT isotherms reported over a wide range of temperatures i.e., (298–680 K). Figure 9 illustrates the performance of MH-PCTpro for predicting PCT isotherms as a function of temperature. Figures 9 (A) and (C) are the experimentally measured PCT isotherms at four different temperatures for Ce0.05Mg0.9Y0.05 and Al0.05Fe0.05La0.17Ni0.73 respectively, whereas the predicted ones for these two compositions

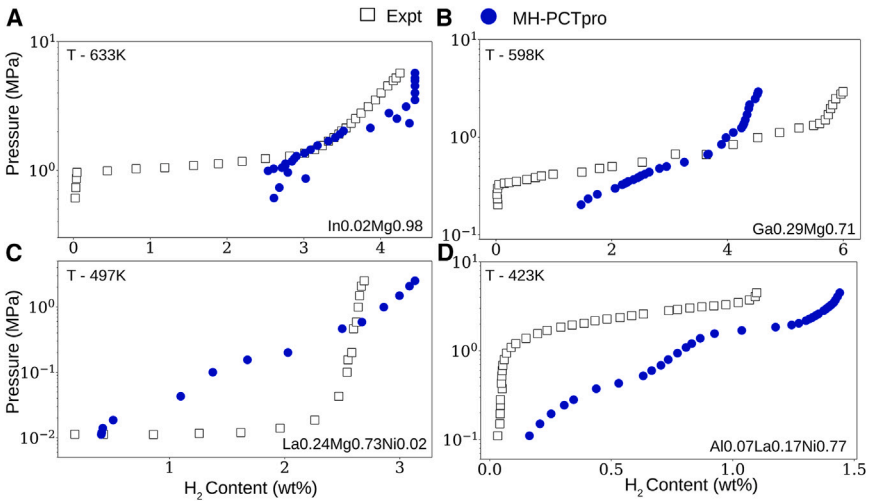


Figure 8. Poorly predicted PCT isotherms for Leave-One-Composition-Out analysis

The experimental data is taken from ref.39–41 and 42 for (A), (B), (C), and (D), respectively.

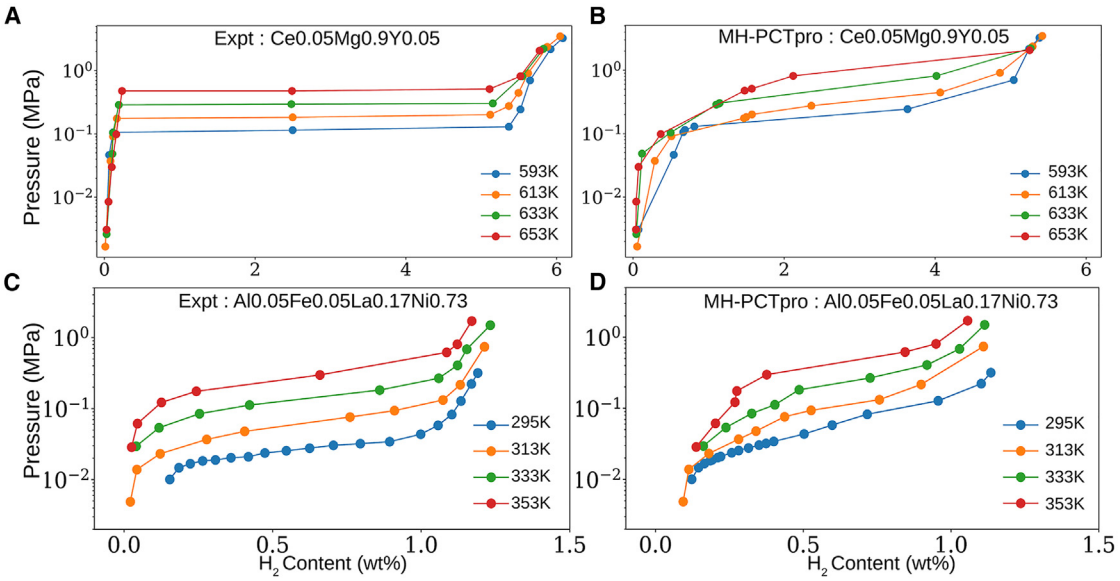


Figure 9. Comparison of experimentally reported and predicted PCT isotherms as function of temperature

(A and C) Illustrate experimentally reported PCT isotherms of Ce0.05Mg0.9Y0.05 and Al0.05Fe0.05La0.17Ni0.73 at four different temperatures, respectively. (B and D) Illustrate predicted PCT isotherms for these two compositions, respectively. The experimental data is taken from 38 for Ce0.05Mg0.9Y0.05 and from 43 for Al0.05Fe0.05La0.17Ni0.73.

are shown in Figures 9 (B) and (D) respectively. The predicted shift in the position of the plateau with temperature is aligned with the experimentally reported values. These predicted PCT isotherms can be mapped to generate a Van't Hoff plot, which leads to the determination of enthalpy and entropy of hydride formation.

Training MH-PCTpro on a specific alloy family

The correctness of predictions depends crucially on the available data. To understand how MH-PCTpro performs under limited data conditions, we conducted an experiment using a specific family of La-based compositions. The original database comprised 237 PCT isotherms of 138 compositions, including 59 PCT isotherms of 31 La-based compositions. We trained MH-PCTpro on the subset of 59 PCT isotherms corresponding to 31 La-based compositions. To evaluate the model's predictability, we employed a Leave-One-Composition-Out (LOCO) strategy. Essentially, we trained MH-PCTpro on 30 of the 31 La-based compositions and then predicted the isotherms of the remaining one composition. This process was repeated for all 31 compositions, enabling the evaluation of the model's accuracy on each one individually. Figure 10 represents the examples from predictions of La-based compositions. The predictions of the model trained on La-based compositions (green circles), closely follows the experimental results (black squares), while the prediction from the general model (blue circles), deviates more. Importantly, the predictions of hydrogen concentration are improved mostly in the lower pressure regions. These improved regions are marked by orange circles to aid an eye. For all 31 La-based compositions, the predictions are comparable and in 40% cases, surpass those of the generalized model trained on a larger dataset of 138 compositions. We also evaluated the performance of MH-PCTpro, trained exclusively on La-based compositions, in predicting isotherms for compositions doped with unseen lanthanides (Ce and Nd). Figure 11 (A) Ce0.03Mg0.89Ni0.08 and (B) Mg0.6Nd0.07Ni0.33 illustrate improved results for unseen compositions, the La-based model (green circles) closely follows or even surpasses the prediction of the generalized model (blue circles), demonstrating an ability to predict for other unseen elements of lanthanide series. Hence, by employing targeted training strategies, we can effectively build predictive models for specific composition families, even with limited data availability.

In conclusion, the developed ML model, MH-PCTpro, exhibits significant promise in predicting PCT isotherms for metal alloys. The MH-PCTpro is trained on simple and easily calculable features incorporating elemental properties, absorption temperature, and hydrogen pressure. Including hydriding properties of elements as features has notably enhanced the model's predictive capabilities. The model has been validated across different alloy families, spanning a wide range of temperature and pressure conditions, and has exhibited good agreement with experimental results. We are listing key features of MH-PCTpro.

MH-PCTpro is trained on 237 PCT isotherms data of 138 compositions comprising more than 25 elements (Al, Ce, Co, Cr, Cu, Fe, Ga, In, La, Li, Mg, Mn, Mo, Nb, Nd, Ni, Pd, Si, Sm, Sn, Ta, Ti, V, Y, Zn, and Zr).

The predictions of PCT isotherms for various classes of metal compositions, encompassing Mg, Li, Al, Si, 3d transition metals, and lanthanides as constituent elements, align well with experimental results.

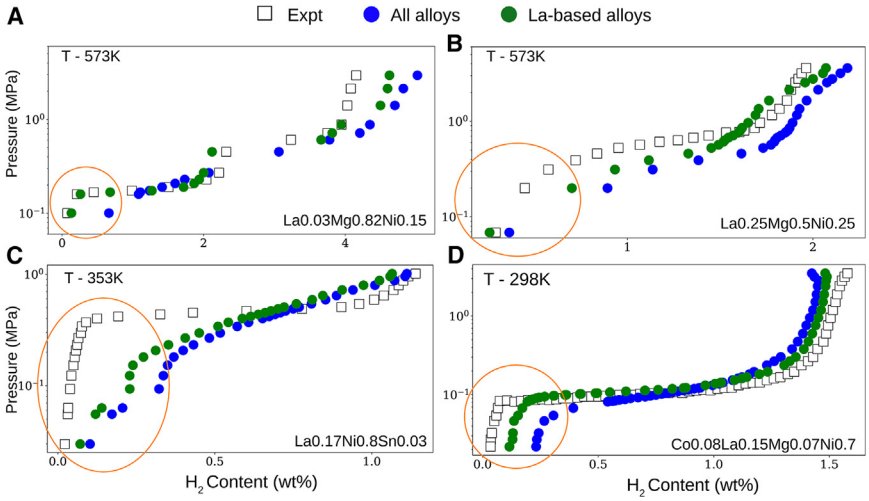


Figure 10. Improved prediction of model trained on La-based compositions

The predictions of the model trained on La-based compositions (green circles) predicts the experimental data better than the general model in most of the cases, demonstrating effective learning with family-specific data. The experimental data is taken from ref.44–46 and 47 for (A), (B), (C), and (D), respectively.

The MH-PCTpro is capable of predicting the alterations in the shape of the PCT isotherms resulting from introducing dopants in the host alloy.

It also predicts the temperature-dependent variations in plateau pressure of compositions. As a result, the predicted PCT isotherms can be mapped onto a Van't Hoff plot to determine the enthalpy and entropy of the hydride formation of an alloy.

MH-PCTpro exhibits limitations especially for compositions consist of elements outside the training set (e.g., non-metals like In and Ga) or PCT isotherms at less-represented temperature regions (400-500K), highlighting the model's dependency on the availability of labeled data.

We also demonstrated that employing targeted family training can be a valuable approach for effectively predicting the PCT isotherms within the family, even for unseen elements, particularly when data are scarce.

Overall, the MH-PCTpro model effectively predicts the PCT isotherms of metal alloys and provides valuable thermodynamic insights. It can be effectively used to curtail the experimental space for the time consuming PCT experiments. The results obtained from this model are promising and we are taking it ahead to integrate MH-PCTpro with experimental workflow. This combined approach holds significant potential for accelerating the discovery of efficient metal hydrides suitable for solid-state hydrogen storage.

Limitations of the study

Despite its overall appreciable performance, MH-PCTpro has certain limitations. Like any ML model, the accuracy depends upon the available data for training the model. A dip in the performance is observed for alloys with less representation in the training set. For example, training data are limited in alloys with non-metals like In and Ga or predictions in the temperature range, such as 400–500K, which is correlated with the observed performance of the model. Another possible front to improve the performance is the feature set. Features capturing the correct chemistry/physics associated with target properties are crucial for the model to excel in predictions. Features or descriptors are crucial for these models, and there is always room for improvement compared to existing ones. We will be working on these two fronts in future.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Kavita Joshi (k.joshi@ncl.res.in).

Materials availability

This study did not generate new materials.

Data and code availability

- We have provided a list of DOIs of the papers used to construct the MH-PCT dataset in Table S6. Due to the bounding terms and conditions of the project, the processed data used for this work will be available after two years, that is after the completion of the project (i.e., March 2027).

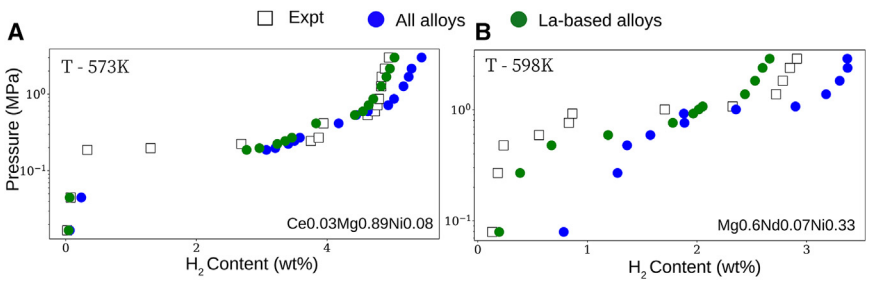


Figure 11. Predicted PCT isotherms of compositions consisting of unseen lanthanide Ce and Nd

The predictions of the model trained on La-based compositions (green circles) demonstrating effective learning with family-specific data. The experimental data is taken from ref.35 for (A) and from ref.48 for (B).

- All the information related to ML model is available in this paper's supplemental information. Due to the bounding terms and conditions of the project, the custom code used for this work will be available after two years, that is after the completion of the project (i.e., March 2027).
- Any additional information required to reanalyze the data reported in this work is available from the lead contact upon request. Until the data and code can be made publicly available, the authors can predict the PCT isotherms for any composition using the models (if its applicable to that composition) and make it available upon request.

ACKNOWLEDGMENTS

CSIR is acknowledged for financial support through grant HCP-44-05 (HEART). Authors also acknowledge SERB-POWER grant (SPG/2022/000890). A.V. thanks DST for INSPIRE fellowship. Authors thank Ameeya Bhushan Sahoo and Nikhil Wilson for their help in assembling the MH-PCT database.

AUTHOR CONTRIBUTIONS

A.V.: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, and visualization. K.J.: Conceptualization, writing – review and editing, visualization, supervision, project administration, and funding acquisition.

DECLARATION OF INTERESTS

We have filed an application for patent based on this work.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES

During the preparation of this work the author(s) used ChatGPT and Grammarly in order to enhance the clarity and readability of the manuscript. After using these tools/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.112251.

Received: May 2, 2024
Revised: August 12, 2024
Accepted: March 17, 2025
Published: March 20, 2025

REFERENCES

1. Abe, J.O., Popoola, A.p.i., Ajenifuja, E., and Popoola, O.M. (2019). Hydrogen energy, economy and storage: Review and recommendation. Int. J. Hydrogen Energy *44*, 15072–15086.
2. Yang, J., Sudik, A., Wolverton, C., and Siegel, D.J. (2010). High capacity hydrogen storage materials: attributes for automotive applications and techniques for materials discovery. Chem. Soc. Rev. *39*, 656–675.
3. Nazir, G., Rehman, A., Hussain, S., Aftab, S., Heo, K., Ikram, M., Patil, S.A., and Aizaz Ud Din, M. (2022). Recent advances and reliable assessment of solid-state materials for hydrogen storage: A step forward toward a sustainable H2 economy. Advanced Sustainable Systems *6*, 2200276.
4. Kumar, A., Muthukumar, P., Sharma, P., and Kumar, E.A. (2022). Absorption based solid state hydrogen storage system: A review. Sustain. Energy Technol. Assessments *52*, 102204.
5. Lai, Q., Sun, Y., Wang, T., Modi, P., Cazorla, C., Demirci, U.B., Ares Fernandez, J.R., Leardini, F., and Aguey-Zinsou, K.-F. (2019). How to design hydrogen storage materials? fundamentals, synthesis, and storage tanks. Advanced Sustainable Systems *3*, 1900043.
6. Barthélémy, H., Weber, M., and Barbier, F. (2017). Hydrogen storage: Recent improvements and industrial perspectives. Int. J. Hydrogen Energy *42*, 7254–7262.
7. Klopčič, N., Grimmer, I., Winkler, F., Sartory, M., and Trattner, A. (2023). A review on metal hydride materials for hydrogen storage. J. Energy Storage *72*, 108456.
8. Gross, K.J., Carrington, K.R., Barcelo, S., Karkamkar, A., Purewal, J., Ma, S., Zhou, H.-C., Dantzer, P., Ott, K., Pivak, Y., et al. (2016). Recommended best practices for the characterization of storage properties of hydrogen storage materials. Tech. Rep. EMN-HYMARC (EMN-HyMARC). National Renewable Energy Lab. (NREL). https://www1.eere.energy.gov/hydrogenandfuelcells/pdfs/best_practices_hydrogen_storage.pdf.
9. (2009). Introduction. In Nanomaterials for Solid State Hydrogen Storage. Fuel Cells and Hydrogen EnergyNanomaterials for Solid State Hydrogen Storage. Fuel Cells and Hydrogen Energy (Boston, MA: Springer US), pp. 1–81, ISBN 978-0-387-77712-2. https://doi.org/10.1007/978-0-387-77712-2_1.
10. Zepon, G., Silva, B.H., Zlotea, C., Botta, W.J., and Champion, Y. (2021). Thermodynamic modelling of hydrogen-multicomponent alloy systems: Calculating pressure-composition-temperature diagrams. Acta Mater. *215*, 117070.
11. Ponsoni, J.B., Aranda, V., Nascimento, T.d.S., Strozi, R.B., Botta, W.J., and Zepon, G. (2022). Design of multicomponent alloys with c14 laves phase structure for hydrogen storage assisted by computational thermodynamic. Acta Mater. *240*, 118317.
12. Agarwal, S., Mehta, S., and Joshi, K. (2020). Understanding the ml black box with simple descriptors to predict cluster–adsorbate interaction energy. New J. Chem. *44*, 8545–8553.
13. Moosavi, S.M., Jablonka, K.M., and Smit, B. (2020). The role of machine learning in the understanding and design of materials. J. Am. Chem. Soc. *142*, 20273–20287.
14. Rangel-Martinez, D., Nigam, K.d.p., and Ricardez-Sandoval, L.A. (2021). Machine learning on sustainable energy: A review and outlook on renewable energy systems, catalysis, smart grid and energy storage. Chem. Eng. Res. Des. *174*, 414–441.
15. Verma, A., Wilson, N., and Joshi, K. (2024). Solid state hydrogen storage: Decoding the path through machine learning. Int. J. Hydrogen Energy *50*, 1518–1528.
16. Lu, Z., Wang, J., Wu, Y., Guo, X., and Xiao, W. (2022). Predicting hydrogen storage capacity of V–Ti–Cr–Fe alloy via ensemble machine learning. Int. J. Hydrogen Energy *47*, 34583–34593.
17. Nations, S., Nandi, T., Ramazani, A., Wang, S., and Duan, Y. (2023). Metal hydride composition-derived parameters as machine learning features for material design and h2 storage. J. Energy Storage *70*, 107980.
18. Batalović, K., Radaković, J., Kuzmanović, B., Medić Ilić, M., and Paskaš Mamula, B. (2023). Machine learning-based high-throughput screening of Mg-containing alloys for hydrogen storage and energy conversion applications. J. Energy Storage *68*, 107720.
19. Zhou, P., Xiao, X., Zhu, X., Chen, Y., Lu, W., Piao, M., Cao, Z., Lu, M., Fang, F., Li, Z., et al. (2023). Machine learning enabled customization of performance-oriented hydrogen storage materials for fuel cell systems. Energy Storage Mater. *63*, 102964.
20. Witman, M., Ling, S., Grant, D.M., Walker, G.S., Agarwal, S., Stavila, V., and Allendorf, M.D. (2020). Extracting an empirical intermetallic hydride

design principle from limited data via interpretable machine learning. J. Phys. Chem. Lett. *11*, 40–47.

21. Guyon, I., and Elisseeff, A. (2003). An introduction to variable and feature selection. J. Mach. Learn. Res. *3*, 1157–1182.
22. Wilson, N., Verma, A., Maharana, P.R., Sahoo, A.B., and Joshi, K. (2024). Hystor: An experimental database of hydrogen storage properties for various metal alloy classes. Int. J. Hydrogen Energy *90*, 460–469.
23. Kim, J.M., Ha, T., Lee, J., Lee, Y.-S., and Shim, J.-H. (2023). Prediction of pressure-composition-temperature curves of AB2-type hydrogen storage alloys by machine learning. Met. Mater. Int. *29*, 861–869.
24. Young, K., Ouchi, T., and Fetcenko, M.a. (2009). Pressure–composition–temperature hysteresis in C14 laves phase alloys: Part 1. simple ternary alloys. J. Alloys Compd. *480*, 428–433.
25. Guéguen, A., Joubert, J.-M., and Latroche, M. (2011). Influence of the C14 Ti35. 4V32. 3Fe32. 3 Laves phase on the hydrogenation properties of the body-centered cubic compound Ti24. 5V59. 3Fe16. 2. J. Alloys Compd. *509*, 3013–3018.
26. Luo, L., Li, Y., Zhai, T., Hu, F., Zhao, Z., Bian, X., and Wu, W. (2019). Microstructure and hydrogen storage properties of V48Fe12Ti15-xCr25Alx (x= 0, 1) alloys. Int. J. Hydrogen Energy *44*, 25188–25198.
27. Charbonnier, V., Enoki, H., Asano, K., Kim, H., and Sakaki, K. (2021). Tuning the hydrogenation properties of ti1+ ycr2-xmnx laves phase compounds for high pressure metal-hydride compressors. Int. J. Hydrogen Energy *46*, 36369–36380.
28. Young, K., Ouchi, T., Liu, Y., Reichman, B., Mays, W., and Fetcenko, M.a. (2009). Structural and electrochemical properties of tixzr7- xni10. J. Alloys Compd. *480*, 521–528.
29. Yong, H., Wang, S., Ma, J., Zhang, K., Zhao, D., Hu, J., and Zhang, Y. (2021). Dual-tuning of de/hydrogenation kinetic properties of mg-based hydrogen storage alloy by building a ni-/co-multi-platform collaborative system. Int. J. Hydrogen Energy *46*, 24202–24213.
30. Shao, H., Chen, C., Liu, T., and Li, X. (2014). Phase, microstructure and hydrogen storage properties of mg–ni materials synthesized from metal nanoparticles. Nanotechnology *25*, 135704.
31. Chen, R.r., Chen, X.y., Ding, X., Li, X.z., Guo, J.j., Ding, H.s., Su, Y.q., and Fu, H.z. (2018). Effects of ti/mn ratio on microstructure and hydrogen storage properties of ti-v-mn alloys. J. Alloys Compd. *748*, 171–178.
32. Dehouche, Z., Savard, M., Laurencelle, F., and Goyette, J. (2005). Ti–v–mn based alloys for hydrogen compression system. J. Alloys Compd. *400*, 276–280.
33. Ponthieu, M., Fernández, J.f., Cuevas, F., Ares, J.r., Leardini, F., Bodega, J., and Sánchez, C. (2013). Reversible hydrogen storage in the Ni-rich pseudo-binary Mg6Pd0. 25Ni0. 75 intermetallic compound: Reaction pathway, thermodynamic and kinetic properties. J. Alloys Compd. *548*, 96–104.
34. Wang, Y., Zhou, Z., Zhou, W., Xu, L., Guo, J., and Lan, Z. (2016). Effects of in-situ formed Mg2Si phase on the hydrogen storage properties of MgLi solid solution alloys. Mater. Des. *111*, 248–252.
35. Meng, J., Wang, X.-L., Chou, K.-C., and Li, Q. (2013). Hydrogen storage properties of graphite-modified Mg-Ni-Ce composites prepared by mechanical milling followed by microwave sintering. Metall. Mater. Trans. A *44*, 58–67.
36. Yong, H., Guo, S., Yuan, Z., Qi, Y., Zhao, D., and Zhang, Y. (2019). Improved hydrogen storage kinetics and thermodynamics of RE-Mg-based alloy by co-doping Ce–Y. Int. J. Hydrogen Energy *44*, 16765–16776.
37. Zhang, Y., Cui, S., Yuan, Z., Gao, J., Dong, X., Qi, Y., and Guo, S. (2018). Hydrogen storage performances of REMg 11 Ni (RE= Sm, Y) alloys prepared by mechanical milling. Metall. Mater. Trans. A *49*, 376–384.
38. Xie, D.h., Li, P., Zeng, C.x., Sun, J.w., and Qu, X.h. (2009). Effect of substitution of nd for mg on the hydrogen storage properties of Mg2Ni alloy. J. Alloys Compd. *478*, 96–102.
39. Zhou, C., Fang, Z.Z., Lu, J., Luo, X., Ren, C., Fan, P., Ren, Y., and Zhang, X. (2014). Thermodynamic destabilization of magnesium hydride using Mg-based solid solution alloys. J. Phys. Chem. C *118*, 11526–11535.
40. Wu, D., Ouyang, L., Wu, C., Wang, H., Liu, J., Sun, L., and Zhu, M. (2015). Phase transition and hydrogen storage properties of Mg–Ga alloy. J. Alloys Compd. *642*, 180–184.
41. Lin, H.j., Ouyang, L.z., Wang, H., Liu, J.w., and Zhu, M. (2012). Phase transition and hydrogen storage properties of melt-spun mg3lani0. 1 alloy. Int. J. Hydrogen Energy *37*, 1145–1150.
42. Choudhari, M.S., and Sharma, V.K. (2021). Performance investigations on hydrogen-based thermochemical energy storage system through finite volume method and thermodynamic simulation. Int. J. Energy Res. *45*, 20156–20175.
43. Kazakov, A.n., Romanov, I.a., Mitrokhin, S.v., and Kiseleva, E.a. (2020). Experimental investigations of AB5-type alloys for hydrogen separation from biological gas streams. Int. J. Hydrogen Energy *45*, 4685–4692.
44. Li, Q., Pan, Y., Leng, H., and Chou, K. (2014). Structures and properties of Mg–La–Ni ternary hydrogen storage alloys by microwave-assisted activation synthesis. Int. J. Hydrogen Energy *39*, 14247–14254.
45. Ouyang, L.z., Yao, L., Dong, H.w., Li, L.q., and Zhu, M. (2009). Hydrogen storage properties of lamg2ni prepared by induction melting. J. Alloys Compd. *485*, 507–509.
46. Dehouche, Z., Grimard, N., Laurencelle, F., Goyette, J., and Bose, T.k. (2005). Hydride alloys properties investigations for hydrogen sorption compressor. J. Alloys Compd. *399*, 224–236.
47. Huang, T., Yuan, X., Yu, J., Wu, Z., Han, J., Sun, G., Xu, N., and Zhang, Y. (2012). Effects of annealing treatment and partial substitution of Cu for Co on phase composition and hydrogen storage performance of La0. 7Mg0. 3Ni3. 2Co0. 35 alloy. Int. J. Hydrogen Energy *37*, 1074–1079.
48. Xie, D.h., Li, P., Zeng, C.x., Sun, J.w., and Qu, X.h. (2009). Effect of substitution of Nd for Mg on the hydrogen storage properties of Mg2Ni alloy. J. Alloys Compd. *478*, 96–102.
49. Rohatgi, A. (2022). Webplotdigitizer: Version 4.6. https://automeris.io/WebPlotDigitizer.
50. Jain, A., Ong, S.P., Hautier, G., Chen, W., Richards, W.D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., and Persson, K.A. (2013). Commentary: The materials project: A materials genome approach to accelerating materials innovation. APL Mater. *1*.
51. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
52. Pandala, S.R. (2022). lazypredict. https://pypi.org/project/lazypredict/pythonlibraryforrapidMLmodelcomparison.
53. Floriano, R., Zepon, G., Edalati, K., Fontana, G.L.B.G., Mohammadi, A., Ma, Z., Li, H.-W., and Contieri, R.J. (2020). Hydrogen storage in TiZrNbFeNi high entropy alloys, designed by thermodynamic calculations. Int. J. Hydrogen Energy *45*, 33759–33770.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
PCT isotherms dataset	This paper	SI
Software and algorithms		
Web PlotDigitizer	Rohatgi[49]	https://automeris.io/WebPlotDigitize/
scikitlearn python library	Pedregosa et al.[50]	https://scikit-learn.org/
LazyPredict	Pandala[51]	https://pypi.org/project/lazypredict/

METHOD DETAILS

To train the ML model, an experimental database of metal hydrides and their absorption PCT plots (MH-PCT) is compiled from published literature spanning the years 1990–2021. The Web PlotDigitizer is used to extract the requisite data points from the PCT plots.[49] During the extraction process, we interpolated additional points on the curves to increase the density of data points. This interpolation helped in creating a more comprehensive dataset. For each extracted data point from PCT plots, we referred back to the original papers to verify the composition and ensure that the reported plots were derived from experimental results. This step is crucial for maintaining the accuracy and reliability of our database. Notably, the units for variables such as pressure, temperature, and hydrogen weight capacity were inconsistent across different papers. Therefore, we standardized these units across all data points. The finalized database consists of 138 unique alloy compositions, each accompanied by their corresponding PCT isotherms measured at various temperatures. DOIs of the papers used to construct the MH-PCT database are provided in Table S6. In total, MH-PCT database comprises more than 14,000 data points from 237 distinct PCT isotherms. These curves are visualized in Figure S5A, covering a wide spectrum of absorption temperatures, hydrogen pressure, and hydrogen concentration ranges. The Figure S5B illustrates occurrence of elements within the database. Notably, the dataset encompasses diverse families of alloys consisting of more than 25 hydriding and non-hydriding elements. However, we do observe the inadequate representation of certain elements, such as Li, Si, In, Ga, Zn, Mo, Pd, Nd, and Sm. Additionally, there are only a few PCT isotherms available in the temperature range of 400-500K, as depicted in Figure S5C. We have analyzed the models' predictions for these under-represented elements and temperature range to gain a better understanding of the model's performance.

For the development of a predictive model, it is crucial to select features that optimize model's performance while also being readily available or easy to compute. The hydrogen storage properties of metal hydrides depend on pressure, temperature, and composition. In our previous ML framework, HEART,[15] we have demonstrated the importance of absorption temperature as a feature for accurate prediction of hydrogen weight capacity of metal alloy. Our work also highlights the significance of features such as metal-metal, metal-hydrogen interactions along with compositional and elemental properties. To train MH-PCTpro, we expanded the feature set of HEART, by incorporating the following additional features: (1) hydrogen absorption pressure and (2) elemental hydriding properties, such as the weighted average enthalpy of hydride formation (E_{hyd}) of constituent elements and the volume difference (Δ Vol) between the element's bulk phase and its hydride phase. We obtained the experimentally reported hydriding properties from the Materials Project.[50] The expanded feature set, encompassing twenty-eight distinct features, is presented in Table S3. However, within this set, several features exhibit multicollinearity with each other, that can adversely impact the model's performance. The Pearson correlation coefficients shown in Figure S6 are computed to identify and select features based on their collinearity. The final set consisting of features with collinearity values below 0.9 is tabulated in Table S3. The model's learning and predictability for different feature sets is reported and discussed in the next section.

The scikitlearn python library[51] is used to implement ML model. For the rapid screening of various ML algorithms, we employed LazyPredict[52] which automates the process of building and evaluating machine learning models. The results of the top 10 performing algorithms are listed in Table S4, including their associated error bars. Among all the evaluated algorithms, extra trees regressor (ETR) a tree-based algorithm demonstrated the lowest error and subsequently selected for model development. GridSearchCV is employed to find the optimal values of hyper-parameters. The final set of hyper-parameters used for training MH-PCTpro is provided in Table S5. The dataset is split into 80-20 ratio for the train and validation set. Mean Absolute Error (MAE) and R2 Score, are used to evaluate the performance of the model.

QUANTIFICATION AND STATISTICAL ANALYSIS

Mean Absolute Error (MAE) and R2 Score, are used to evaluate the performance of the model. Details of all statistical analyses can be found above in the results and discussion section.