

Predicting the Hydrogen Release Ability of LiBH₄-based Mixtures by Ensemble Machine Learning

Zhao Ding ^{a,b}, Zhiqian Chen ^{c*},

Tianyi Ma ^d, Chang-Tien Lu ^c, Wenhui Ma ^a, Leon Shaw ^b

^a *Faculty of Metallurgical and Energy Engineering, Kunming University of Science and Technology, Kunming, Yunnan, China, 650093*

^b *Department of Mechanical, Materials and Aerospace Engineering, Illinois Institute of Technology, Chicago, Illinois, USA, 60616*

^c *Department of Computer Science, Virginia Tech, Falls Church, Virginia, USA, 22043*

^d *Discipline of Chemistry, School of Environmental & Life Sciences, Faculty of Science, The University of Newcastle, University Drive, Callaghan, Australia NSW 2308*

Acknowledge

This work was supported under the U.S. National Science Foundation (NSF) with the Award No. CMMI-1261782.

* **Corresponding author:** czq@vt.edu

Predicting the Hydrogen Release Ability of LiBH₄-based Mixtures by Ensemble Machine Learning

Abstract

The prediction of hydrogen release ability is indispensable to evaluating hydrogen storage performance of LiBH₄-based mixtures before experimentation. To achieve this goal, ensemble machine learning is employed to automatically infer the relationship between factors (i.e., sample preparation, mixing conditions and operational variables) and target (H₂ release amount), providing exceptional insight into hydrogen release ability. Specifically, the importance ranking of major variables for the hydrogen release of LiBH₄ has been proposed for the first time based on the constructed uni-component catalysts database. We train our developed EoE model on 2,071 uni-component catalysts data and attempt to predict the hydrogen release amounts of LiBH₄ doping with the unseen bi-component catalysts. The appealing results demonstrate the effectiveness and robustness of EoE. The procedure established in this study presents a novel approach for accelerating the research and development of hydrogen storage materials over various catalysts.

Keywords: Prediction, LiBH₄, Catalysts, Hydrogen Release Ability, Ensemble Machine Learning

1. Introduction

The growing concern with the traditional energy crisis and pollution has gradually led to the replacement of the conventional fossil fuel-based energy structure by renewable energy based structure. Over the last two decades, studies on the renewable hydrogen fuel source, which has also been termed as “Ultimate Power Source” in the 21st century, are rising like mushrooms [1, 2, 3, 4, 5, 6, 7, 8]. Just under this background, tremendous efforts have been devoted to tuning the stable thermodynamics and sluggish kinetics of LiBH₄, which has the greatest chance to serve as on-board hydrogen storage alternative, because of its high gravimetric and volumetric hydrogen densities of 18.5 wt.% and 121 kg H₂/m³, respectively [9].

LiBH₄ + MgH₂ as a promising hydrogen storage system, has the ability to reversibly release and uptake ~ 5.0 wt.% H₂ at 265 °C, which is the highest reversible H₂ storage capacity ever reported for the LiBH₄ + MgH₂ system in solid-state [10]. This superior hydrogen storage capacity is achieved through high-energy ball milling of MgH₂ at ambient temperature along with aerosol spraying of LiBH₄ dissolved in tetrahydrofuran solution [11]. LiBH₄ + NH₂NH₂ as another encouraging hydrogen storage materials, is capable to release ~ 13.0 wt.% H₂ at temperature as low as 140 °C [12]. However, its application is limited by the tough and complex re-hydrogenation, due to the final dehydrogenation product of Li₃BN₂ and BN [13]. LiBH₄ + Li₂NH has been shown to hydrogenate at 37 °C under 50 bar of H₂, which has been attribute to the formation of a new phase between Li₂NH and LiBH₄ after ball milling for more than 24 h or heating to 147 °C [14]. It should be noted that 37 °C is the lowest operating temperature for lithium amide/imide systems to date [14]. Another latest effort is through nanoconfinement of the CO₂-activated carbon aerogel; the dehydrogenation temperature from LiBH₄ + Ca(BH₄)₂ has been reduced by

93 °C when compared with that of bulk $\text{LiBH}_4 + \text{Ca}(\text{BH}_4)_2$ [15]. Furthermore, the apparent activation energies has been decreased from 204 kJ/mol for the bulk to 130 kJ/mol for the nanoconfined mixtures [15]. However, none of the hydrogen storage materials studied so far has met the performance targets for on-board hydrogen storage system for light-duty fuel cell vehicles established by U.S. Department of Energy, which demands the storage materials release ~ 0.0065 kg H_2 /kg system with ~ 1500 operation cycles life at the temperature of $45 \sim 85$ °C [16]. In order to hit the target, researchers [17, 18, 19] have spent much time developing and synthesizing various catalysts to improve the hydrogen release performance of LiBH_4 . However, whether the novel catalysts are worthy of in-depth investigation and whether it is possible to meet the researchers' expectation are difficult to predict. Therefore, effective and efficient prediction on the hydrogen release ability before the experimentation is very necessary for boosting research of hydrogen storage materials.

As the data accumulates, many patterns are hidden in massive variables, and human can hardly handle such a large amount of data. The modern machine learning tool allows the properties of materials to be pre-estimated before evaluation in an actual laboratory experiment. High-throughput computational pre-estimation has become prevalent, enabling researchers to compute the properties of tens of thousands of materials. There is an increasing infrastructure of machine learning algorithms for tuning computational models. Machine learning techniques are widely-recognized for handling complicated problems of large combinatorial spaces or non-linearity, while conventional methods cannot solve within a reasonable computational cost. Recently, machine learning is reported to have powerfulness and potential in molecule and material modelling, such as searching for superior lithium ion conducting solids [20], predicting atomic-scale properties [21], rational computation-guided co-design [22], high-throughput synthesis [23], and addressing challenges in molecular and materials research [24].

However, predicting the hydrogen release ability is still non-trivial due to several facts: **(I) No clear patterns have been discovered between material properties and its hydrogen release ability.** All the related material properties, such as its intrinsic nature and experimental environment configuration, are indispensable to determine the hydrogen release ability. Therefore, it is impossible to identify a simple or analytical form between these properties and hydrogen release amount. **(II) Manual selection among models and parameters is inevitable and tedious.** To achieve the best performance for the practical application of machine learning on predicting hydrogen release ability, one needs to fully understand the characteristics of different models, and make a reasonable decision on parameters, according to different data scenarios, which is a time-consuming process. **(III) Optimization on different datasets for a single model is challenging.** Each single machine learning model has its pros and cons, depending on the data. Therefore, a single model can hardly handle all possible cases, even with carefully parameter tuning.

In this paper, we study ensemble learning methodology for predicting the hydrogen release ability, which integrates multiple machine learning models into one model to boost predictive performance, markedly improving the error level and stability without manually tuning. Inspired by ensemble methodology, we ensemble a set of ensemble machine learning models, namely EoE (ensemble of an ensemble), to further improve and robustify prediction performance. The contributions of this paper are summarized as follows:

- The importance analysis is performed on the hydrogen storage properties of LiBH_4 based on the constructed uni-component catalysts database.
- The accuracy of prediction on the hydrogen release ability via EoE is compared with the main-stream machine learning techniques.
- EoE is attempted to evaluate the hydrogen ability of LiBH_4 doping with bi-component catalysts which are not included in the uni-component catalysts database.

2. Database Construction

H_2 release data from light metal hydrides and complex hydrides published from 2003 to 2018 are collected to establish the database. In the beginning, all the possible features that may be critical to determine the desorption performance are obtained after reviewing more than 100 publications. During the re-examination, only the papers about the catalytic effects of different catalysts, which includes graphene [25], Mg_3La and TiCl_3 [26], $\text{La}_2\text{Mg}_{17}$ [27], Ce_2S_3 [28], SrH_2 [29], SrF_2 [30], CaF_2 [31], metal oxides (V_2O_3 , SnO_2 or ZrO_2) [32, 33], metal halides (TiCl_3 , TiF_3 , ZnF_2 , MgF_2 , MgCl_2 , CaCl_2 , SrCl_2 or FeCl_3) [34], carbon nanotubes [35], AlF_3 [36], MnF_2 and MnCl_2 [37], SiO_2 or TiF_3 [38], MoS_2 [39], TiF_3 [40], TiO_2 [41], NbF_5 [42], various oxides (Fe_2O_3 , V_2O_5 , Nb_2O_5 , TiO_2 , SiO_2) [43], metal (Mg, Al, Ti, V, Cr, or Sc) or metal hydride (MgH_2 , TiH_2 or CaH_2) [44], different types of carbon (graphite, purified single-walled carbon nanotubes and activated carbon composites) [45], on the hydrogen release from LiBH_4 are selected.

In order to ensure the applicability of the model, we have already ruled out the rare catalysts with special requirements, such as carbon-supported Pd nanoparticles [46], carbon-supported Pt nanoparticles, [47] graphene supported Pt nanoparticles [48] and Ru nanoparticles supported on multiwalled carbon nanotubes [49], while constructing the database, even doping these unique catalysts would result in promising results. It is also worth noting that, we only consider uni-component catalysts at first, the composite catalysts (binary or ternary component catalysts) are susceptible to considerable uncertainties because of unavoidable and/or unaccounted reactions between the catalysts themselves usually have a powerful impact on the catalytic effects of hydrogen release ability, which could jeopardize the accuracy of the prediction. Therefore, our database is made up of the 2071 data points extracted from 32 publications, the schematic diagram of the uni-component catalysts database for the hydrogen storage properties of LiBH_4 is shown in Figure 1.

The database is generated on the basis of 14 input variables, all of which are regarded as being either continuous (such as dehydrogenation temperature) or categorical (such as catalysts) according to their characteristics. We have divided these data into three groups, sample preparation (catalysts, molar ratio of LiBH_4 to catalyst), mixing conditions (hand milling time, ball milling time, ball-to-materials ratio, ball milling atmosphere) and operational variables (dehydrogenation temperature, heating rate, dehydrogenation atmosphere, holding time, re-hydrogenation temperature, re-hydrogenation time, re-hydrogenation atmosphere and the corresponding cycles). The ranges of the corresponding input variables are summarized in Table 1. Here, we should note that, the hydrogen release amount, $\text{m}\% \text{H}_2$, serves as the sole output (performance) variable in the modelling.

3. Model

100 This section firstly introduces the basic process of machine learning and then presents the limitation of prevail linear models. Finally, details of ensemble learning are elaborated for predicting the hydrogen release ability.

As expressed in Figure 2, our dataset is randomly divided into two sets, i.e., the training set and testing set, with the ratio of 7:3. **This random split is repeated 10 times, and we calculate the mean and variance of the performance metrics, which is equivalent to cross-validation or N-fold split. Performance under 10-fold split has been listed in Figure S1. Noted the related metric definition will be introduced at section 4.1.** The testing set is for evaluating how precise the model can predict the hydrogen release ability given properties, and therefore, the testing set is assumed to be invisible during the training process [50, 51]. The machine learning model will run on the training dataset to derive latent patterns
110 between all the input variables and hydrogen release amount, and then the model will be evaluated on testing dataset. If the performance on the testing dataset is good enough, then we assume that it will also have a good performance on any other unseen dataset.

Linear Models are widely recognized that they suffer from several shortcomings: (i) limited to linear relationships between variable and target, (ii) only focus on the mean of the dependent variable about the target, (iii) sensitive to outliers. Therefore, applying linear models on a non-linear dataset is insufficient. To circumvent these issues and reveal non-linear patterns between variables and hydrogen release ability, we propose to employ ensemble learning, which significantly outperforms traditional approaches by integrating multiple machine learning models. Ensemble methods utilize multiple machine learning algorithms to obtain better predictive performance than that of any of single algorithm. Ensemble learning consists of a set of learning models, and automatically learn the weights of each model to construct a robust ensemble. In practice, ensemble machine learning will have better predictive performance, especially when there is diversity among the models [52]. Figure 3 shows the schematic of ensemble machine learning. Specifically, we apply several state-of-the-art models implemented by scikit-learn [53]:
120

- *Decision Tree* [54] is a decision-making tool that uses a binary tree and its possible combinatorial
125 results.
- *Random Forest* [55] is a typical ensemble learning model that operates by building a set of decision trees and yielding average prediction of a separate tree. Random decision forests is superior to decision trees since it can solve the over-fitting issue, since random forests contain a number of trees.
- *AdaBoost* [56] is an estimator that fits on the dataset and then fits extra copies of the regressor on the
130 same data, but the weights of instances are adjusted according to current prediction performance. As such, subsequent regressors focus more on tough cases.
- *Bagging* [57] is an ensemble fitting on random subsets of the original dataset, and then make a final decision based on aggregated prediction. Bagging method is to used to robustify the original set of
135 models, by introducing randomness during the training process and then ensemble their predictions.
- *Extra Trees* [58] implements a meta estimator that fits several random decision trees on different

sub-samples of the dataset and utilizes the mean of trees to boost the predictive performance and reduce the variance.

- *Gradient Boosting Regression Tree* (GBRT) considers extra models:

$$F(x) = \sum_{m=1}^M \gamma_m h_m(x),$$

140 where $h_m(x)$ is the function called weak models, and γ_m is corresponding weight. GBRT also a decision trees based method, which can handle data of mixed and complex patterns.

- *Histogram-based Gradient Boosting Regression Tree* (Hist) [59] is much faster than Gradient Boosting Regressor for big datasets. The input data is pre-processed into integer-valued bins, which dramatically reduces the number of splitting points.

Inspired by ensemble learning, we ensemble the ensemble models above as 2-layer ensemble (i.e., an 145 ensemble of ensemble, or EoE model) to further improve the accuracy of the predictive model. With integrating a set of different machine learning models and obtaining the average predicted values, EoE has been proved to be capable of automatically selecting the best constitute model and robustifying the performance. In order to balance out their variances and therefore robustify the performance, various ensemble models are employed to build the home-developed EoE learning (See Figure 3).

150 4. Results and Discussion

4.1. Feature Dependence Analysis

Feature analysis is an important step in machine learning. In many applications, researchers hope to gain insight by analyzing how a model can predict a target and what features it uses. Therefore, before designing a predictive model, we perform importance analysis to obtain insight of contribution for each 155 variable. Generally, importance rating is a normalized score indicating how critical a feature is for the building of the model. A higher score for an attribute means higher relative importance. **Specifically, the state of the art tree-based models, i.e., Gradient-Boosted Regression Trees (GBRT), Decision tree, Ada Decision Tree and Random Forest are all employed to extract the importance scores.** When a variable splits a node, the Gini impurity of the two descendent child nodes decreases compared with the parent 160 node. The total Gini decreases of all variables equal to the importance. Gini importance calculates each variable importance as the sum of the splits, proportional to the number of samples it splits. The partial dependence plot shows if there exist a linear relationship between the target and a variable, allowing variables to be compared in terms of importance. **Based on the above-mentioned models, the importance ranking of the top eight features are plotted in Figure 4.**

165 Not surprisingly, dehydrogenation temperature is the most important variable for the H₂ desorption from LiBH₄. As generally accepted, there are several elementary steps for the complete hydrogen release of LiBH₄ through the reaction $2 \text{LiBH}_4 \rightarrow 2 \text{LiH} + 2 \text{B} + 3 \text{H}_2$ [60]. Dehydrogenation mechanism differs from different temperature range, together with different amounts of hydrogen release. Heating bulk LiBH₄ to 105 ~ 112 °C, the structure would transform from orthorhombic, which is more stable at room 170 temperature, to hexagonal, accompanying with ~ 0.1 wt.% H₂ release at its first desorption peak. The following 0.5 ~ 1.0 wt.% H₂ release occurs with the melting of LiBH₄ (275 ~ 278 °C). The largest

desorption with more than 9.0 wt.% H₂ release happens at temperature ranges from 400 to 680 °C. The last hydrogen desorption is the decomposition of LiH, requires temperature as high as ~ 827 °C, and thus is barely mentioned at all. Moreover, LiH cannot be rehydrided into LiBH₄ until the temperature is higher than 600 °C under 35 MPa of hydrogen. Therefore, LiBH₄ would have become a qualified on-board storage alternative for hydrogen vehicles if the largest hydrogen desorption could happen at a lower temperature, the very problem we need to face before the application of LiBH₄, is the high-temperature stability.

The thermal stability of LiBH₄ is dependent to a considerable degree on the electronegativity of the central Li atom [61]. Because of this, various catalysts have been tried to lower the dehydriding temperature by substituting Li atoms with the other metal atoms with higher electronegativity. Typically, with doping TiCl₃, the cation exchange reaction $3\text{LiBH}_4 + \text{TiCl}_3 \rightarrow 3\text{LiCl} + \text{Ti}(\text{BH}_4)_3$ has already taken place during ball milling, and thus the Ti(BH₄)₃ product starts the hydrogen release at room temperature through reaction $\text{Ti}(\text{BH}_4)_3 \rightarrow \text{TiH}_2 + 2.5\text{B} + 0.25\text{B}_2\text{H}_6 + 4.25\text{H}_2$ [34]. Similarly, the ball-milled LiBH₄ + TiF₃ mixture has the ability to release 5.0 wt.% and 6.4 wt.% hydrogen at the temperature of 250 °C and 500 °C, respectively, according to the reaction $3\text{LiBH}_4 + \text{TiF}_3 \rightarrow 3\text{LiF} + \text{TiB}_2 + \text{B} + 6\text{H}_2$ [40]. Metal oxides (MO_x) have also been proved to be effective to reduce the dehydrogenation temperature of LiBH₄, as a result of the representative redox reaction such as $\text{LiBH}_4 + \text{MO}_x \rightarrow \text{LiMO}_x + \text{B} + 2\text{H}_2$ [43]. For example, the ball-milled LiBH₄ + Fe₂O₃ mixture is shown to release 6.0 wt.% hydrogen at temperature $\leq 200^\circ\text{C}$, and the onset temperature of dehydrogenation is ~ 100 °C [43]. According to the above mentioned discussions, it is reasonable to obtain the catalyst as the secondary important factor for hydrogen release ability. Noted that this ranking is aggregated by summing effects of all the uni-component catalysts over our database.

It is discovered that the initial dehydrogenation temperature of LiBH₄ has reduced from ~100 °C for LiBH₄ + 0.1 TiX₃ (X=Cl or F) to room temperature (~ 25 °C) for LiBH₄ + 0.5 TiX₃ [34], which means the amounts of the catalyst would definitely influence the hydrogen release ability. However, it should also note that the decay of the hydrogen storage capacity of LiBH₄ + 0.5 TiX₃ mixtures during cycles of dehydrogenation and re-hydrogenation are more serious than that of LiBH₄ + 0.1 TiX₃, because the boron loss along with diborane emission is unrecoverable [34]. Furthermore, due to the higher hydrogen density of LiBH₄ when compared to the catalyst, the mass fraction of hydrogen release would be lower with doping more catalysts. Therefore, as an important factor, it is very essential to test out the optimum doping amount of the corresponding catalyst, which could balance the reaction temperature and storage reversibility.

Detailed studies [62, 63, 64, 65] on LiBH₄-based reactive hydride composites have shown that the hydrogen release reaction mechanism depends not only on temperature, but also on the dehydrogenation atmosphere. Taking the LiBH₄ + YH₃ composite for instance [65], the incubation period of dehydriding reaction would be obviously shortened with the increasing hydrogen or argon back pressure from 0.4 to 0.7 MPa, which shows a great accelerating effect on hydrogen release rate. Only negligible amount of hydrogen can be released under hydrogen pressure of 1.0 MPa but the hydrogen release amount under argon pressure of 1.0 MPa, remains much the same [65]. The analyses of the dehydrogenation behaviors of the LiBH₄ + MgH₂ mixtures by in-situ XRD combined with pressure observation [62], indicates that hydrogen release from LiBH₄ would be suppressed at the beginning under hydrogen pressure of 5 bar,

while simultaneous decomposition of LiBH_4 and MgH_2 could be observed under hydrogen pressure ≤ 3 bar. In addition, it is believed that the formation of $\text{Li}_2\text{B}_{12}\text{H}_{12}$, which has been regarded as a “dead end” of the dehydrogenating reaction of LiBH_4 , could be efficiently suppressed by increasing the gas back pressure [63]. Therefore, other than temperature, the dehydrogenation atmosphere (Ar and/or H_2 gas back pressure) also plays a decisive role for the decomposition pathways of LiBH_4 . According to Figure 4, the absolute scores of dehydrogenation atmosphere and molar ratio of LiBH_4 to catalyst obtained from different models are very close, as a result, they are tied as the third important variables for the hydrogen release ability.

The other two important variables are ball milling time and ball-to-materials ratio. The rate-limiting step for reactions between solid reactants is very likely to be nucleation and growth of the products. For example, the solid-state dehydrogenation of $\text{LiBH}_4 + \text{MgH}_2$ mixture at $265\text{ }^\circ\text{C}$ is proved to be controlled by the nucleation/growth of the LiH and MgB_2 at the interfacial area between MgH_2 and LiBH_4 [66, 67]. In this case, longer ball milling time and greater ball-to-materials ratio could offer larger interfacial areas between the solid reactants by producing particles with finer size, which would expedite the hydrogen release. Moreover, this would also help generate a finer and more uniform particle size system, which should be very effective to avoid/reduce the gradual decay in the reversible hydrogen storage capacity during cycles of dehydrogenation and re-hydrogenation, by reducing the tendency for a particular particle to grow at the expense of other particles [68]. Therefore, it is safe to conclude that both ball milling time and ball-to-materials ratio contribute to the reactions kinetics of hydrogen release.

Heating rate, as the joint-last important feature, is used as a common input variable in related fields, which has important effects on the hydrogen release temperature. The most notorious example is, the slower heating rate, the lower hydrogen desorption peak temperature would be observed during DSC measurements of $\text{LiBH}_4 + \text{MgH}_2$ mixture [67]. Thus, the optimum heating rate should be the fastest rate that within the instrumental sensitivity, which would also ensure enough time to complete the reaction [69].

We have further evaluated correlations among all variables and found that there is little linearity between any pair of variables. The relationship among the top 1-4 and top 5-8 important features (in Figure 4) are illustrated respectively in Figure 5 and Figure 6 with 4×4 sub-figures respectively.

Figure 5 and 6 are the kernel pairplots of variables based on training dataset, predicting the distribution of unseen data. The diagonal sub-figures show the distribution of variables where the x-axis defines the range of the specific variable as indicated, while the y-axis reflects the distribution density of the variable. Data points and lines with red, green and blue indicate high, middle and low levels of hydrogen release amount respectively. For instance, the first sub-figure in the first row of Figure 5 shows the distribution of dehydrogenation temperature, which can distinguish different hydrogen release amounts. However, different levels of hydrogen release amount are almost overlapped for regarding any of the other variables shown in the diagonal sub-figures of Figure 5 and 6, which means that they are not as significant as dehydrogenation temperature in estimating hydrogen release amount. The second to fourth sub-figures in first row of Figure 5 show dehydrogenation temperature (y-axis) as a function of dehydrogenation H_2 pressure, dehydrogenation air pressure, and heating rate (x-axis) respectively. Similarly, the y-axis and x-axis of the sub-figures in other rows are marked clearly. For each non-diagonal sub-figure, we have performed a linear model that is shown as lines throughout the data points. The shadowed areas indicate

the locations with the highest probability for unseen data. As shown in Figure 5 and 6, linear model
255 does not fit the patterns properly in all non-diagonal sub-figures due to the non-linearity in the pair-wise
relationship. Therefore, it is safe to conclude that, the relationship among the most important features
are beyond linearity and thereby, an accurate non-linear model is required. We have marked points red,
green and blue, which are employed to represent low, middle and high level of hydrogen release amounts,
respectively. Furthermore, in order to make this more persuasive, the complete relationship among top
260 1-8 important features is also attached in Figure S2 in the Supplementary Materials.

4.2. Model Accuracy

To demonstrate the effectiveness of the proposed EoE, several competitive linear predictive models are
compared: Linear Regression, Ridge Regression and Lasso Regression. Multi-layer Perceptron (MLP)
regressor, as a typical neural networks, is also used as a strong baseline. In the experiments, we use
265 several metrics to evaluate error and correlation w.r.t. experimental data. Specifically, Mean Squared
Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are used to calculate
errors, while R^2 Explained Variance (EV), Spearman (SM) and Pearson (PS) are employed to measure
correlation level. Generally speaking, high error or low correlation means inferior performance. Detailed
interpretation of the metric can be found in the Supplementary Materials.

270 In the evaluation, MSE is selected as the major error metric, and R^2 as major correlation metric. As
shown in Table 2, LASSO models performed the worst, which has MSE of 6.158 and R^2 score is 0.403.
MLP and the other linear models (Linear and Ridge) also achieve relatively high MSE (beyond 3) and
low R^2 (below 0.7). While the ensemble models (i.e., Gradient Boosting, Decision Tree, Ada Decision
Tree, Bagging, Extra Tree, Random Forest and Histogram based Gradient Boosting) improve both MSE
275 and R^2 of linear models significantly, achieving MSE of below 2 and R^2 of beyond 0.8. These results
imply that ensemble models generally accurately characterize the pattern between variables and target,
compared with MLP and linear models. EoE methods outperformed all the other baselines, and have
robust performance since it integrates all ensemble model. Specifically, EoE achieves the highest R^2
score (0.888) and lowest MSE (1.144). This shows that EoE successfully integrates all sub-models and
280 implicitly selects the best model, yielding outstanding and robust prediction performance. Other error
(MAE, RMSE) and correlation (Explained Variance(EV), Spearman (SM) and Pearson (PS)) metrics
also justify the advantage of EoE, showing that EoE is still the best model in prediction task on uni-
component catalyst dataset. Figure 7 visualizes error of each model, where each sub-figure shows the
relationship between experimental value (y axis) and predicted value obtained via different models (x
285 axis) mentioned in Table 2. Errors of ensemble models (in purple) concentrate around perfect prediction
standard ($y=x$), compared with linear model and MLP (in blue). EoE (in yellow) improves ensemble
methods marginally by at most 0.68. The difference between predicted values and experimental results,
namely residues, of each model is visualized in Figure 8, where each sub-figure shows the relationship
between predicted value (x axis) and residue (y axis) obtained via different models. Linear models and
290 MLP have larger residues around $y=0$ than ensemble models and EoE, particularly in the region when
predicted hydrogen release volume is large. This trend means that residue increases for linear models
and MLP as predicted hydrogen release volume increases. On the contrary, ensemble models and EoE
do not have such behaviour, performing stably on both large or small value. Note that MAE of EoE is

not the best, since MAE treat large and small error with the same weight, but MSE/RMSE punish more
295 on large error.

4.3. Application to Bi-component Catalyst Database

One interesting question should be raised here “how well the EoE model predict the hydrogen release ability of LiBH_4 doping with bi-component catalysts that are not included in our previous database?” As mentioned previously, the specific database is only exclusive for uni-component catalysts. It is well
300 accepted that the activation energy of dehydrating reaction is effectively reduced by doping the uni-component catalyst (metal oxides, halides and hydrides), which has been reflected by the much lower initial hydrogen release temperature when compared to that of prime LiBH_4 . However, the catalytic effect of a uni-component catalyst is limited and in most of the cases, a uni-component catalyst shows little effect on the reversibility of hydrogen release. As a result, many researchers have tried to add
305 bi-component catalysts into LiBH_4 , expecting that the other component would either further improve the catalytic effect of the hydrogen desorption, or catalyze the reverse hydrogen absorption.

Assume that the trained model would be able to extrapolate the catalytic effects of bi-component catalysts on the hydrogen release from LiBH_4 into a reasonable value of hydrogen release amount from the catalytic performance of uni-component catalysts over the database. For this test, 236 data points
310 on the dehydrogenation properties of LiBH_4 with doping different bi-component catalysts are collected to construct a typical database of “bi-component catalyst”. The 236 data points are also from the cited 32 papers but they are isolated from building the “uni-component catalyst” database. After training the twelve models listed in Table 2 on uni-component catalyst database, this time the bi-component catalysts database is selected to be the testing set to evaluate the corresponding models by calculating error and
315 residue level. It is worth noting that, two typical lines are employed to enable estimating the accuracy of models by visual inspection, data points present on the identity line ($y=x$) indicating the predicted value is exactly equal to real experimental result, while best fit line shows a linear line learned from the predicted data. Because the distance between the identity line and best fit line is much closer than that of the other conventional models (not shown here) in comparison, we can conclude that the prediction
320 via EoE is the most reliable. Specifically, Figure 9 (a) presents that the prediction error for EoE achieves a high R^2 of 0.834 w.r.t. identity line, which means the accuracy of EoE is 83.4%. Figure 9 (b) shows the residues in training and test data distribution with a high R^2 values of 0.889 and 0.834 w.r.t. with zero residue line ($y=0$), respectively, which justifies that EoE is stable and not overfitting across different datasets. These results demonstrate the developed EoE model is effective not only for the prediction of
325 hydrogen release ability of LiBH_4 with uni-catalyst catalysts, but also for that of bi-component catalysts, motivating our further prediction on multi-component catalysts.

Before concluding, several examples of its application are given in details. It has been reported that with the addition of AlF_3 , hydrogen release at the temperature ~ 100 °C becomes possible due to the reaction between LiBH_4 and AlF_3 , such as $3 \text{LiBH}_4 + 2 \text{AlF}_3 \longrightarrow \text{Li}_3\text{AlF}_6 + \text{Al} + 3 \text{B} + 6 \text{H}_2$ [36]. Then
330 Li_3AlF_6 would react with LiBH_4 with a hydrogen desorption peak temperature at ~ 300 °C through $\text{Li}_3\text{AlF}_6 + 3 \text{LiBH}_4 \longrightarrow 6 \text{LiF} + \text{Al} + 3 \text{B} + 6 \text{H}_2$. However, the formation of LiF and B products from these two elementary dehydrating reactions, could be regarded as “dead end” of the hydrogen release, because the direct rehydrogenation of LiF and B are extremely difficult or even impossible under the

proposed experimental conditions (450 °C, 9.2 MPa, 24 h), as shown in Figure 10, the hydrogen release
335 of $3\text{LiBH}_4 + \text{AlF}_3$ during the 2nd dehydrogenation is ~ 0.5 wt.%. Yuan [36] believed that adding TiF_3
would improve the hydrogen re-adsorption of dehydrided products and thus modify the reversibility of
 $3\text{LiBH}_4 + \text{AlF}_3$ system, just like the catalytic effect of Ti-base compound on the hydrogen release and
uptake of NaAlH_4 [70, 71]. EoE is employed to predict the hydrogen release ability of $3\text{LiBH}_4 + \text{AlF}_3$
+ 0.2TiF_3 mixtures after training on the hydrogen release properties of $\text{LiBH}_4 + \text{AlF}_3$ and $\text{LiBH}_4 +$
340 TiF_3 mixtures under the same experimental conditions. Green dots in Figure 10 denote the predication
of hydrogen release amounts w.r.t. temperature during the first dehydrogenation, while the blue dots
represent the prediction of the second dehydrogenation.

The $\text{Li}_2\text{S} + \text{CeH}_3$ mixtures is also included in the bi-component catalysts database. The existence of
 Li_2S is able to facilitate the cleavage of the ionic bond between B and H, and thus efficiently decrease the
345 required energy for the decomposition of LiBH_4 [39]. Inspired by Shim’s research [63], Wang [28] tried to
add CeH_3 into $\text{LiBH}_4 + \text{Li}_2\text{S}$, anticipating that, CeH_3 could serve as heterogeneous nucleation mediums
during dehydrogenation, just like CeB_6 , and thereby would definitely further prompt the dehydrogenation
kinetic of LiBH_4 . Similarly, the hydrogen release ability of $\text{LiBH}_4 + 10$ wt.% $\text{Li}_2\text{S} + 10$ wt.% CeH_3
mixtures is also predicted via EoE, and the training set includes the hydrogen release performance of
350 $\text{LiBH}_4 + \text{Li}_2\text{S}$ and $\text{LiBH}_4 + \text{CeH}_3$ mixtures. Blue dots in Figure 11 stand for the prediction of the
hydrogen release amounts w.r.t. dehydrogenation time at 350 °C.

The lines connecting all these predicted dots are drawn by interpolation algorithm. As shown in Figure
10, with temperature increasing, the predicted amounts of hydrogen release (green line) and experiment
results (yellow triangles) during the first dehydrogenation show a similar tendency, while the predicted
355 values during the second dehydrogenation (blue line) is almost overlapped with experiment data (yellow
rectangles). The results indicate that even though the dehydrogenation kinetics of $3\text{LiBH}_4 + \text{AlF}_3$ has
been ameliorated by the addition of TiF_3 , the expected effect on the reversibility is limited, only ~ 0.1
wt.% more hydrogen is released from $3\text{LiBH}_4 + \text{AlF}_3 + 0.2\text{TiF}_3$ mixtures than that from $3\text{LiBH}_4 +$
 AlF_3 mixtures during the second dehydrogenation [36]. Figure 11 exhibits that the foreseen hydrogen
360 release performance is almost exactly what happened in the experiment, confirming that, hydrogen release
becomes faster with the introduction of CeH_3 into $\text{LiBH}_4 + \text{Li}_2\text{S}$ mixtures [28].

The results are attractive, which corroborates the practical application of our EoE for the main tar-
get of this study: accelerating the research and development of hydrogen storage materials over various
catalysts. Based on the analysis above, the developed EoE is able to make relatively accurate predictions
365 of the hydrogen release ability from LiBH_4 doping with uni- and/or bi-component catalysts at a dis-
cretionary ratio over the database within seconds. Theoretically, the prediction are more accurate with
more and more perfect relevant catalysts database. In fact, as the pre-experimental guide, the overall
relative residue for the prediction via EoE is completely acceptable. However, EoE is still a black-box
model which lacks of interpretability, and always predict with implicit reasons. We plan to design an
370 interpretable model that can extract meaningful and explicit rules for explaining the prediction behavior,
in order to make the whole process more persuasive.

5. Conclusion

In this study, we have demonstrated, for the first time, the importance ranking of major variables in the hydrogen release ability from LiBH_4 is temperature > catalyst > atmosphere \approx catalysts amount > 375 holding time > ball-milling time > heating rate \approx ball to materials ratio via training Gradient Boosting Regression Trees, **Random Forest**, **Ada Decision Tree** and **Decision Tree** on the uni-component catalysts database. We have developed a novel model termed as ensemble of ensemble machine learning (EoE) to make a prediction of hydrogen release of LiBH_4 over various catalysts before experimentation. For the prediction on the uni-component catalysts doping, the home-developed EoE has achieved the low- 380 est error (MSE: 1.144) and highest correlation (R^2 : 0.888) among the main-stream machine learning models. We also illustrate the application of EoE by testing the model on 236 unseen data points of the hydrogen release performance from LiBH_4 doping with bi-component catalysts, after training on the uni-component catalysts with 2,071 data points. The deviation between the predicted and experimental results is reasonable, and does not entirely discount the study. Therefore, we consider the prediction 385 qualified for the pre-experimental guide. Furthermore, the prediction would be more accurate, with the growth of data in volume and dimensionality. To the best of our knowledge, this is the first time the ensemble machine learning has been employed to predict the hydrogen release ability of LiBH_4 -based hydrogen storage materials.

Moreover, the developed procedure is easily adaptable to predicting the hydrogen release ability of 390 other hydrogen storage materials. This will greatly accelerate the accelerating the research and development of hydrogen storage materials, achieve earlier and better the requirement for on-board hydrogen storage system for light-duty fuel cell.

References

- [1] P. Chen, Z. Xiong, J. Luo, J. Lin, K. L. Tan, Interaction of hydrogen with metal nitrides and imides, 395 Nature 420 (6913) (2002) 302.
- [2] A. Dillon, K. Jones, T. Bekkedahl, C. Kiang, D. Bethune, M. Heben, Storage of hydrogen in single-walled carbon nanotubes, Nature 386 (6623) (1997) 377.
- [3] L. Schlapbach, Technology: Hydrogen-fuelled vehicles, Nature 460 (7257) (2009) 809.
- [4] J. F. Hull, Y. Himeda, W.-H. Wang, B. Hashiguchi, R. Periana, D. J. Szalda, J. T. Muckerman, 400 E. Fujita, Reversible hydrogen storage using CO_2 and a proton-switchable iridium catalyst in aqueous media under mild temperatures and pressures, Nature Chem. 4 (5) (2012) 383.
- [5] G. Li, H. Kobayashi, J. M. Taylor, R. Ikeda, Y. Kubota, K. Kato, M. Takata, T. Yamamoto, S. Toh, S. Matsumura, Hydrogen storage in pd nanocrystals covered with a metal-organic framework, Nature Mater. 13 (8) (2014) 802.
- 405 [6] Z. Xiong, C. K. Yong, G. Wu, P. Chen, W. Shaw, A. Karkamkar, T. Autrey, M. O. Jones, S. R. Johnson, P. P. Edwards, High-capacity hydrogen storage in lithium and sodium amidoboranes, Nature Mater. 7 (2) (2008) 138.

- [7] C. Liu, Y. Fan, M. Liu, H. Cong, H. Cheng, M. S. Dresselhaus, Hydrogen storage in single-walled carbon nanotubes at room temperature, *Science* 286 (5442) (1999) 1127–1129.
- 410 [8] N. L. Rosi, J. Eckert, M. Eddaoudi, D. T. Vodak, Hydrogen storage in microporous metal-organic frameworks, *Science* 300 (5622) (2003) 1127–1127.
- [9] L. Schlapbach, A. Züttel, Hydrogen-storage materials for mobile applications, in: *Materials for sustainable energy: a collection of peer-reviewed research and review articles from nature publishing group*, World Scientific, 2011, pp. 265–270.
- 415 [10] Z. Ding, Y. Lu, L. Li, L. Shaw, High reversible capacity hydrogen storage through nano-libh4+nano-mgh2 system, *Energy Storage Materials* 20 (2019) 24–35.
- [11] Z. Ding, X. Zhao, L. Shaw, Reaction between LiBH₄ and MgH₂ induced by high-energy ball milling, *J. Power Sources* 293 (2015) 236–245.
- [12] T. He, H. Wu, G. Wu, J. Wang, W. Zhou, Z. Xiong, J. Chen, T. Zhang, P. Chen, Borohydride
420 hydrazinates: high hydrogen content materials for hydrogen storage, *Energ. Environ. Sci.* 5 (2) (2012) 5686–5689.
- [13] F. E. Pinkerton, G. P. Meisner, M. S. Meyer, M. P. Balogh, M. D. Kundrat, Hydrogen desorption exceeding ten weight percent from the new quaternary hydride Li₃BN₂H₈, *J. Phys. Chem. B* 109 (1) (2005) 6–8.
- 425 [14] H. Wang, H. Cao, W. Zhang, J. Chen, H. Wu, C. Pistidda, X. Ju, W. Zhou, G. Wu, M. Etter, Li₂NH-LiBH₄: a complex hydride with near ambient hydrogen adsorption and fast lithium ion conduction, *J. Phys. Chem. B* 24 (6) (2018) 1342–1347.
- [15] P. Javadian, D. A. Sheppard, C. E. Buckley, T. R. Jensen, Hydrogen storage properties of nanoconfined LiBH₄-Ca(BH₄)₂, *Nano Energy* 11 (2015) 96–103.
- 430 [16] U. Drive, Target explanation document: onboard hydrogen storage for light-duty fuel cell vehicles (2015).
- [17] H. Li, K. Shin, G. Henkelman, Effects of ensembles, ligand, and strain on adsorbate binding to alloy surfaces, *The Journal of chemical physics* 149 (17) (2018) 174705.
- [18] H. Li, W. Chai, G. Henkelman, Selectivity for ethanol partial oxidation: The unique chemistry of
435 single-atom alloy catalysts on au, ag, and cu (111), *Journal of Materials Chemistry A*.
- [19] H. Li, S. Guo, K. Shin, M. S. Wong, G. Henkelman, Design of a pd-au nitrite reduction catalyst by identifying and optimizing active ensembles, *ACS Catalysis* 9 (9) (2019) 7957–7966.
- [20] K. Fujimura, A. Seko, Y. Koyama, A. Kuwabara, I. Kishida, K. Shitara, C. A. Fisher, H. Moriwake, I. Tanaka, Accelerated materials design of lithium superionic conductors based on first-principles
440 calculations and machine learning algorithms, *Adv. Energy Mater.* 3 (8) (2013) 980–985.
- [21] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, M. Ceriotti, Machine learning unifies the modeling of materials and molecules, *Sci. Adv.* 3 (12) (2017) e1701816.

- [22] A. Mannodi-Kanakkithodi, G. M. Treich, T. D. Huan, R. Ma, M. Tefferi, Y. Cao, G. A. Sotzing, R. Ramprasad, Rational co-design of polymer dielectrics for energy storage, *Adv. Mater.* 28 (30) (2016) 6277–6291.
- [23] P. Raccuglia, K. C. Elbert, P. D. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist, Machine-learning-assisted materials discovery using failed experiments, *Nature* 533 (7601) (2016) 73.
- [24] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, *Nature* 559 (7715) (2018) 547.
- [25] J. Xu, R. Meng, J. Cao, X. Gu, Z. Qi, W. Wang, Z. Chen, Enhanced dehydrogenation and rehydrogenation properties of LiBH_4 catalyzed by graphene, *Int. J. Hydrog. Energy* 38 (6) (2013) 2796–2803.
- [26] T. Sun, H. Wang, Q. Zhang, D. Sun, X. Yao, M. Zhu, Synergetic effects of hydrogenated Mg_3La and TiCl_3 on the dehydrogenation of LiBH_4 , *Journal of Materials Chemistry* 21 (25) (2011) 9179–9184.
- [27] Y. Zhou, Y. Liu, W. Wu, Y. Zhang, M. Gao, H. Pan, Improved hydrogen storage properties of LiBH_4 destabilized by in situ formation of MgH_2 and LaH_3 , *J. Phys. Chem. C* 116 (1) (2011) 1588–1595.
- [28] J. Wang, Z. Wang, Y. Li, D. Ke, X. Lin, S. Han, M. Ma, Effect of nano-sized Ce_2S_3 on reversible hydrogen storage properties of LiBH_4 , *Int. J. Hydrog. Energy* 41 (30) (2016) 13156–13162.
- [29] D. Liu, W. Huang, T. Si, Q. Zhang, Hydrogen storage properties of LiBH_4 destabilized by SrH_2 , *J. Alloys Compd.* 551 (2013) 8–11.
- [30] S. Zhao, C. Wang, D. Liu, Q. Tan, Y. Li, T. Si, Destabilization of LiBH_4 by SrF_2 for reversible hydrogen storage, *Int. J. Hydrog. Energy* 43 (10) (2018) 5098–5103.
- [31] P. P. Yuan, B. H. Liu, B. J. Zhang, Z. P. Li, Reversible hydrogen storage composite based on $6\text{LiBH}_4 + \text{CaF}_2$, *J. Phys. Chem. C* 115 (14) (2011) 7067–7075.
- [32] M. Au, A. Jurgensen, Modified lithium borohydrides for reversible hydrogen storage, *J. Phys. Chem. B* 110 (13) (2006) 7062–7067.
- [33] M. Au, A. Jurgensen, K. Zeigler, Modified lithium borohydrides for reversible hydrogen storage (2), *J. Phys. Chem. B* 110 (51) (2006) 26482–26487.
- [34] M. Au, A. R. Jurgensen, W. A. Spencer, D. L. Anton, F. E. Pinkerton, S.-J. Hwang, C. Kim, R. C. Bowman Jr, Stability and reversibility of lithium borohydrides doped by metal halides and hydrides, *J. Phys. Chem. C* 112 (47) (2008) 18661–18671.
- [35] X. Yu, Z. Wu, Q. Chen, Z. Li, B. Weng, T. Huang, Improved hydrogen storage properties of LiBH_4 destabilized by carbon, *Appl. Phys. Lett.* 90 (3) (2007) 034106.
- [36] P. P. Yuan, B. H. Liu, H. P. Zhu, W. Y. Pan, Z. P. Li, Destabilized dehydrogenation reaction of LiBH_4 by AlF_3 , *J. Alloys Compd.* 557 (2013) 124–129.

- [37] F. Fang, Y. Li, Y. Song, D. Sun, Q. Zhang, L. Ouyang, M. Zhu, Superior destabilization effects of MnF_2 over MnCl_2 in the decomposition of LiBH_4 , *J. Phys. Chem. C* 115 (27) (2011) 13528–13533.
- [38] Y. Zhang, W.-S. Zhang, M.-Q. Fan, S.-S. Liu, H.-L. Chu, Y.-H. Zhang, X.-Y. Gao, L.-X. Sun, Enhanced hydrogen storage performance of LiBH_4 - SiO_2 - TiF_3 composite, *J. Phys. Chem. C* 112 (10) (2008) 4005–4010.
- [39] D. Liang, S. Han, J. Wang, W. Zhang, X. Zhao, Z. Zhao, Effect of MoS_2 on hydrogenation storage properties of LiBH_4 , *J. Solid State Chem.* 211 (2014) 21–24.
- [40] Y. Guo, X. Yu, L. Gao, G. Xia, Z. Guo, H.-K. Liu, Significantly improved dehydrogenation of LiBH_4 destabilized by TiF_3 , *Energ. Environ. Sci.* 3 (4) (2010) 464–469.
- [41] X. Yu, D. Grant, G. Walker, Low-temperature dehydrogenation of LiBH_4 through destabilization with TiO_2 , *J. Phys. Chem. C* 112 (29) (2008) 11059–11062.
- [42] H. Kou, G. Sang, Y. Zhou, X. Wang, Z. Huang, W. Luo, L. Chen, X. Xiao, G. Yang, C. Hu, Enhanced hydrogen storage properties of LiBH_4 modified by NbF_5 , *Int. J. Hydrog. Energy* 39 (22) (2014) 11675–11682.
- [43] X. Yu, D. Grant, G. Walker, Dehydrogenation of LiBH_4 destabilized with various oxides, *J. Phys. Chem. C* 113 (41) (2009) 17945–17949.
- [44] J. Yang, A. Sudik, C. Wolverton, Destabilizing LiBH_4 with a metal ($m = \text{mg, al, ti, v, cr, or sc}$) or metal hydride ($\text{MH}_2 = \text{MgH}_2, \text{TiH}_2, \text{or CaH}_2$), *J. Phys. Chem. C* 111 (51) (2007) 19134–19140.
- [45] Z.-Z. Fang, X.-D. Kang, P. Wang, Improved hydrogen storage properties of LiBH_4 by mechanical milling with various carbon additives, *Int. J. Hydrog. Energy* 35 (15) (2010) 8247–8252.
- [46] J. Xu, X. Yu, J. Ni, Z. Zou, Z. Li, H. Yang, Enhanced catalytic dehydrogenation of LiBH_4 by carbon-supported Pd nanoparticles, *Dalton Trans.* (39) (2009) 8386–8391.
- [47] J. Xu, J. Cao, X. Yu, Z. Zou, D. L. Akins, H. Yang, Enhanced catalytic hydrogen release of LiBH_4 by carbon-supported Pt nanoparticles, *J. Alloys Compd.* 490 (1-2) (2010) 88–92.
- [48] J. Xu, Z. Qi, J. Cao, R. Meng, X. Gu, W. Wang, Z. Chen, Reversible hydrogen desorption from LiBH_4 catalyzed by graphene supported Pt nanoparticles, *Dalton Trans.* 42 (36) (2013) 12926–12933.
- [49] J. Mao, Z. Guo, X. Yu, H. Liu, Enhanced hydrogen sorption properties in the LiBH_4 - MgH_2 system catalysed by Ru nanoparticles supported on multiwalled carbon nanotubes, *J. Alloys Compd.* 509 (15) (2011) 5012–5016.
- [50] H. Li, Z. Zhang, Mining the intrinsic trends of CO_2 solubility in blended solutions, *Journal of CO_2 Utilization* 26 (2018) 496–502.
- [51] H. Li, Z. Zhang, Z.-Z. Zhao, Data-mining for processes in chemistry, materials, and engineering, *Processes* 7 (3) (2019) 151.
- [52] D. Opitz, R. Maclin, Popular ensemble methods: An empirical study, *J. Artif. Intell. Res.* 11 (1999) 169–198.

- [53] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- 515 [54] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, Classification and regression trees. belmont, ca: Wadsworth, International Group (1984) 432.
- [55] T. K. Ho, Random decision forests, in: Proceedings of 3rd international conference on document analysis and recognition, Vol. 1, IEEE, 1995, pp. 278–282.
- [56] Y. Freund, R. E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, *J. Comput. Syst. Sci.* 55 (1) (1997) 119–139.
- 520 [57] G. Louppe, P. Geurts, Ensembles on random patches, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2012, pp. 346–361.
- [58] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, *Mach. Learn.* 63 (1) (2006) 3–42.
- [59] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, Lightgbm: A highly efficient gradient boosting decision tree, in: Advances in Neural Information Processing Systems, 2017, pp. 3146–3154.
- 525 [60] Z. Ding, S. Li, Y. Zhou, Z. Chen, F. Zhang, P. Wu, W. Ma, Libh4 for hydrogen storage - new perspectives, *Nano Mater. Sci. In Press*.doi:10.1016/j.nanoms.2019.09.003.
- [61] Y. Nakamori, K. Miwa, A. Ninomiya, H. Li, N. Ohba, S.-i. Towata, A. Züttel, S.-i. Orimo, Correlation between thermodynamical stabilities of metal borohydrides and cation electronegativites: First-principles calculations and experiments, *Phys. Rev. B* 74 (4) (2006) 045126.
- 530 [62] U. BÅúsenberg, D. B. Ravnsbæk, H. Hagemann, V. DÅžAnna, C. B. Minella, C. Pistidda, W. Van Beek, T. R. Jensen, R. Bormann, M. Dornheim, Pressure and temperature influence on the desorption pathway of the LiBH₄- MgH₂ composite system, *J. Phys. Chem. C* 114 (35) (2010) 15212–15217.
- 535 [63] J.-H. Shim, J.-H. Lim, S.-u. Rather, Y.-S. Lee, D. Reed, Y. Kim, D. Book, Y. W. Cho, Effect of hydrogen back pressure on dehydrogenation behavior of LiBH₄-based reactive hydride composites, *J. Phys. Chem. Lett.* 1 (1) (2009) 59–63.
- [64] Y. Yan, A. Remhof, S.-J. Hwang, H.-W. Li, P. Mauron, S.-i. Orimo, A. Züttel, Pressure and temperature dependence of the decomposition pathway of LiBH₄, *Phys. Chem. Chem. Phys.* 14 (18) (2012) 6514–6519.
- 540 [65] K.-B. Kim, J.-H. Shim, Y. W. Cho, K. H. Oh, Pressure-enhanced dehydrogenation reaction of the LiBH₄-YH₃ composite, *Chem. Comm.* 47 (35) (2011) 9831–9833.
- [66] Z. Ding, P. Wu, L. Shaw, Solid-state hydrogen desorption of 2 mgh₂+ libh₄ nano-mixture: A kinetics mechanism study, *J. Alloys Compd.* 806 (25) (2019) 350–360.
- 545

- [67] Z. Ding, L. Shaw, Enhancement of hydrogen desorption from nano-composite prepared by ball-milling MgH_2 with in-situ aerosol-spraying LiBH_4 , *ACS Sustain. Chem. Eng.* 7 (17) (2019) 15064–15072.
- [68] Z. Ding, H. Li, L. Shaw, New insights into the solid-state hydrogen storage of nanostructured LiBH_4 - MgH_2 system, *Chem. Eng. J.* (under review).
- 550 [69] S. Z. Cheng, *Handbook of thermal analysis and calorimetry: applications to polymers and plastics*, Vol. 3, Elsevier, 2002.
- [70] B. Bogdanović, M. Schwickardi, Ti-doped MgH_2 as a hydrogen-storage material—preparation by Ti-catalyzed hydrogenation of aluminum powder in conjunction with sodium hydride, *Appl. Phys. A* 72 (2) (2001) 221–223.
- 555 [71] X. Xiao, L. Chen, X. Fan, X. Wang, C. Chen, Y. Lei, Q. Wang, Direct synthesis of nanocrystalline MgH_2 complex hydride for hydrogen storage, *Appl. Phys. Lett.* 94 (4) (2009) 041907.
- [72] S. A. Glantz, B. K. Slinker, T. B. Neilands, *Primer of applied regression and analysis of variance*, Vol. 309, McGraw-Hill New York, 1990.
- [73] J. T. Kent, Information gain and a general measure of correlation, *Biometrika* 70 (1) (1983) 163–173.

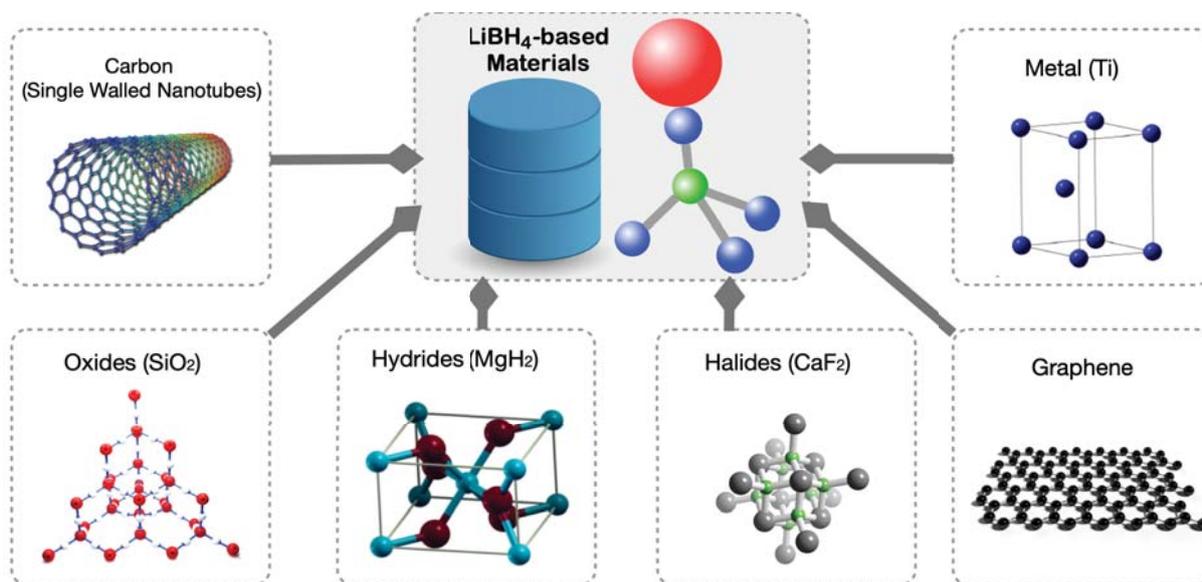


Figure 1: The schematic diagram of uni-component catalysts database for the hydrogen release properties of LiBH_4 .

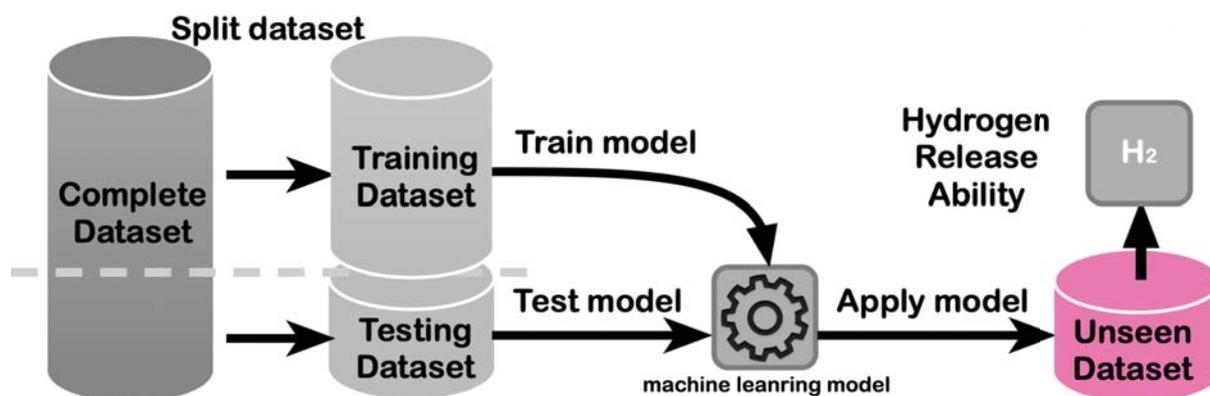


Figure 2: The operation flow-chart of normal machine learning.

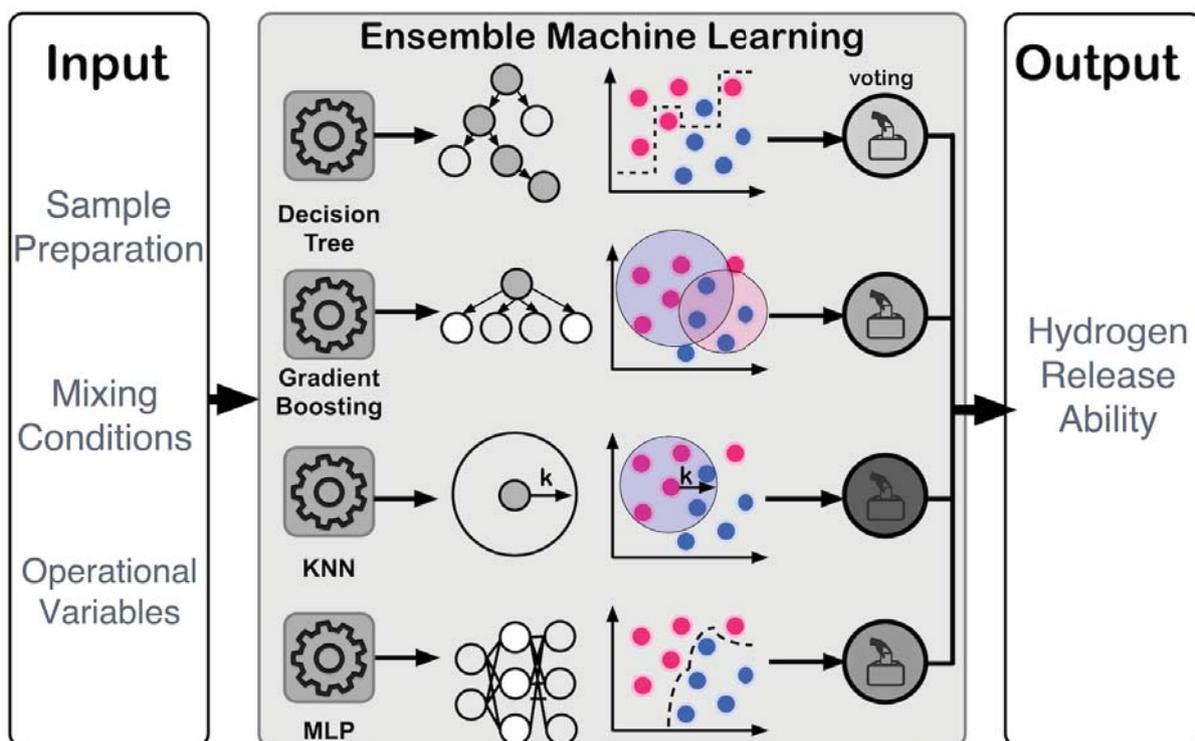


Figure 3: The schematic of home-developed ensemble of ensemble machine learning.

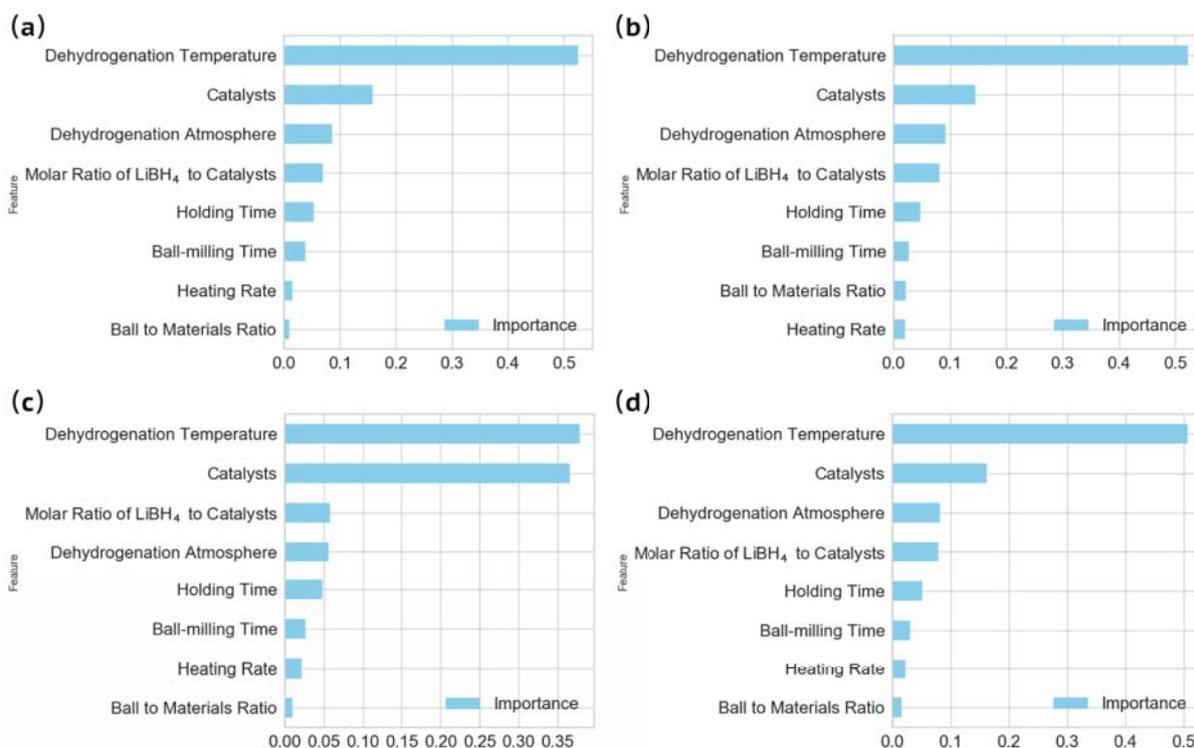


Figure 4: Relative importance ranking among variables with regards to the target (hydrogen release amount) obtained from (a) Gradient Boosting Regression Trees; (b) Decision Tree; (c) Ada Decision Tree; (d) Random Forest.

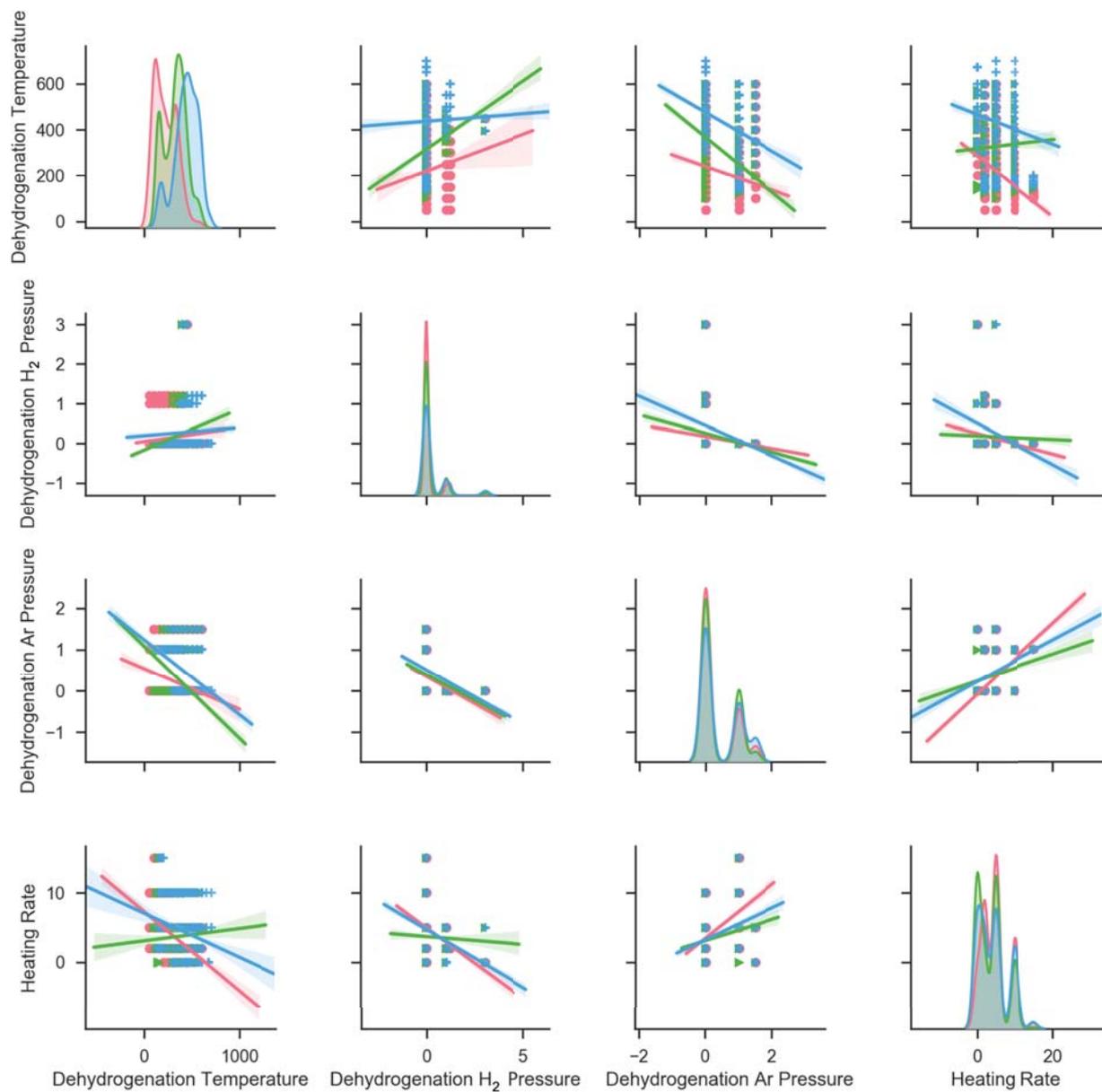


Figure 5: The correlations between top 1-4 variables. Red, green and blue dots indicate instances with low, middle and high level of hydrogen release amount respectively. Red, green and blue lines are generated by linear regression model on the corresponding dots with the same color.

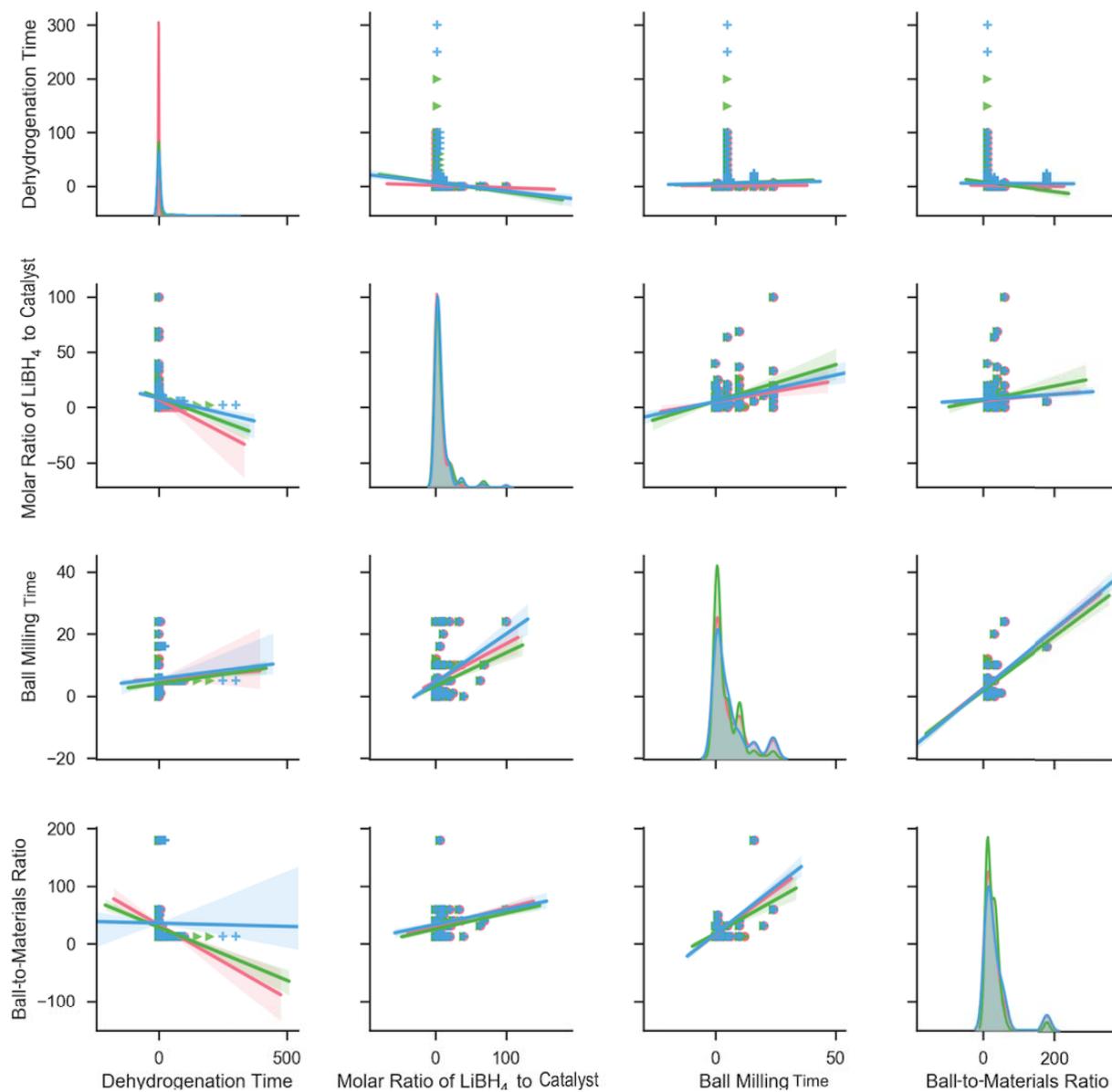


Figure 6: The correlations between top 5-8 variables. Red, green and blue dots indicate instances with low, middle and high level of hydrogen release amount respectively. Red, green and blue lines are generated by linear regression model on the corresponding dots with the same color.

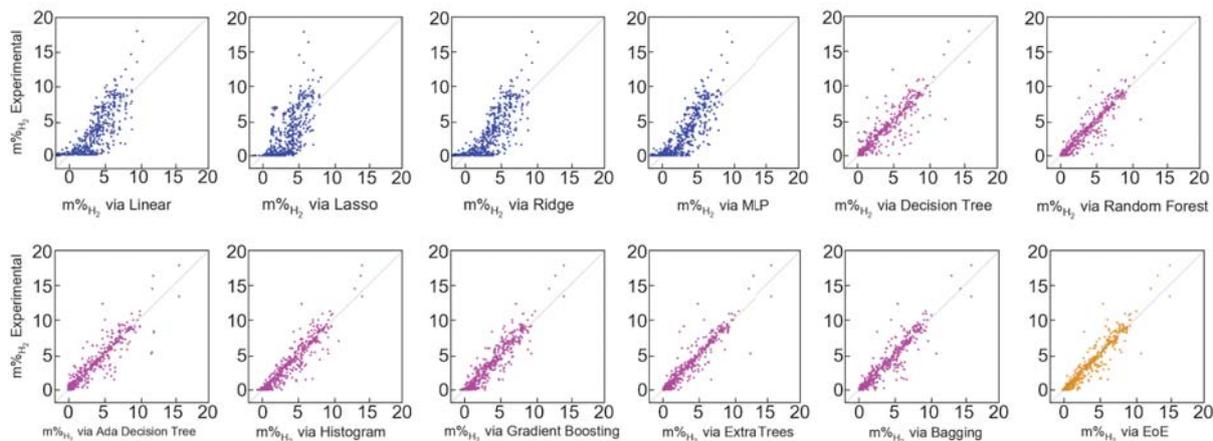


Figure 7: Error comparison of training on uni-component catalysts database for $m\%_{H_2}$. Each sub-figure shows the relationship between experimental results (y axis) and predicted values obtained via the corresponding models (x axis). Perfect prediction line ($y=x$) is shown for comparison.

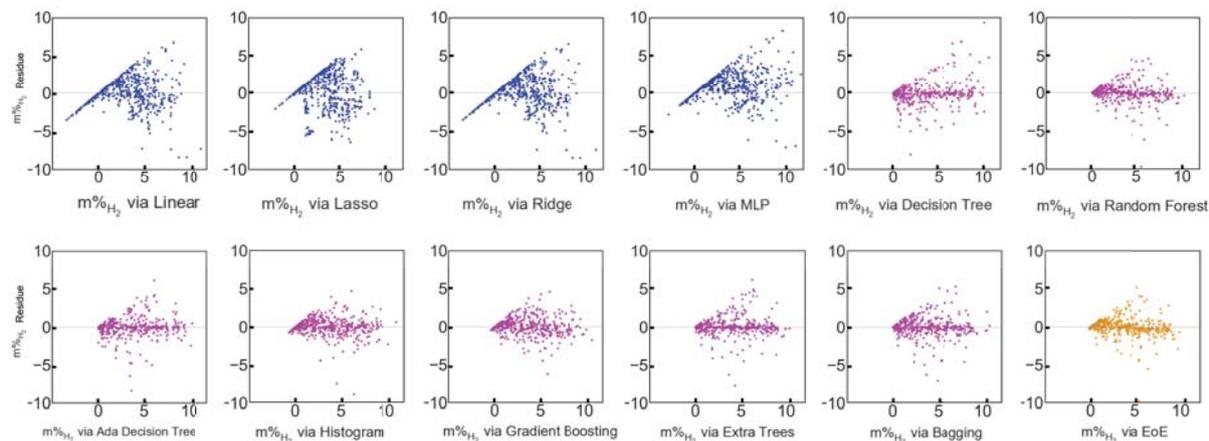


Figure 8: Residue (difference between predicted values and experimental results) comparison of training on uni-component catalysts database for $m\%_{H_2}$. Each sub-figure shows the relationship between residues (y axis) and predicted values obtained via the corresponding models (x axis). Zero residue line ($y=0$) is shown for comparison.

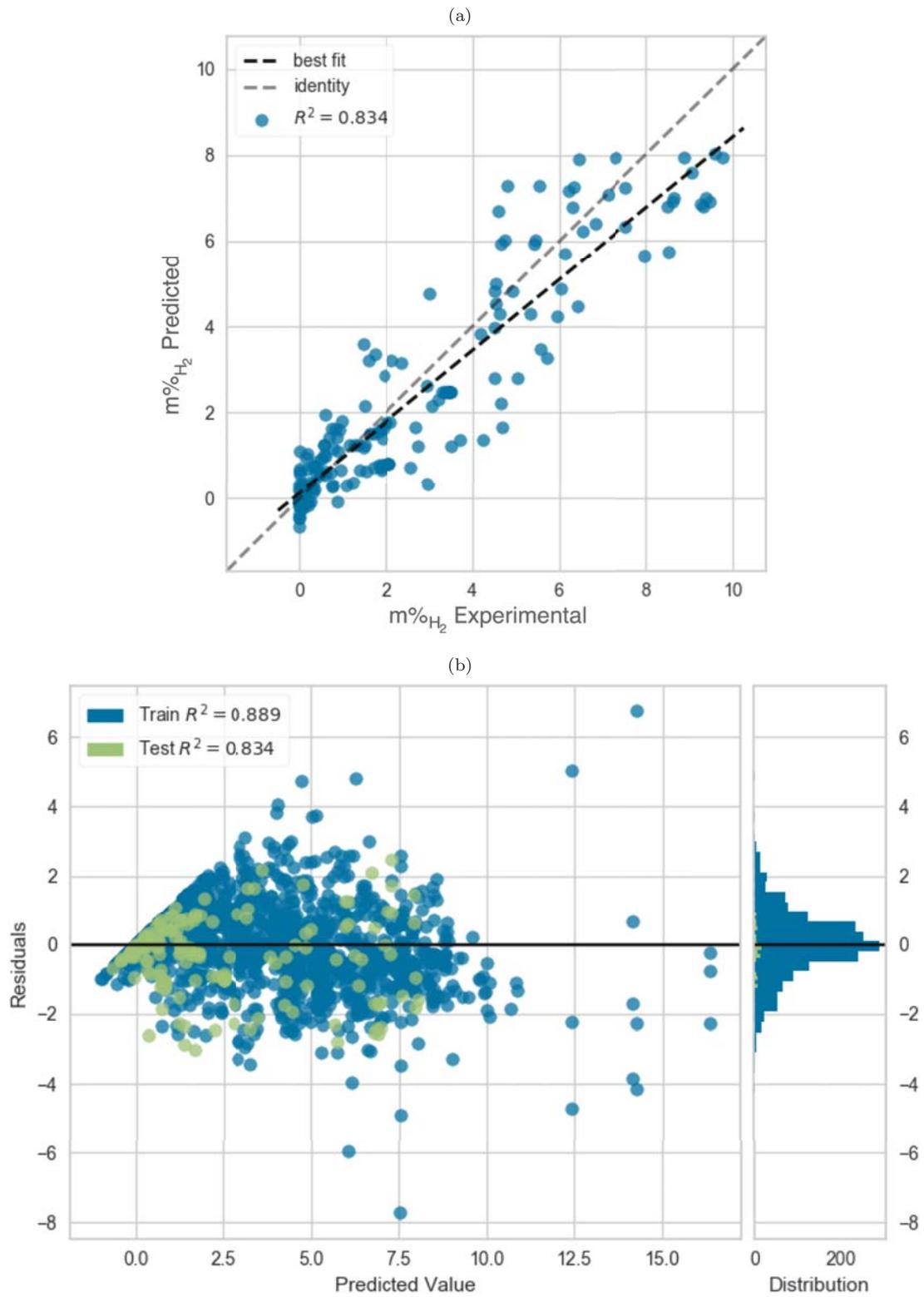


Figure 9: The prediction Error (a) and Residuals (b) for the hydrogen release amount from LiBH_4 doping with bi-component catalysts via EoE.

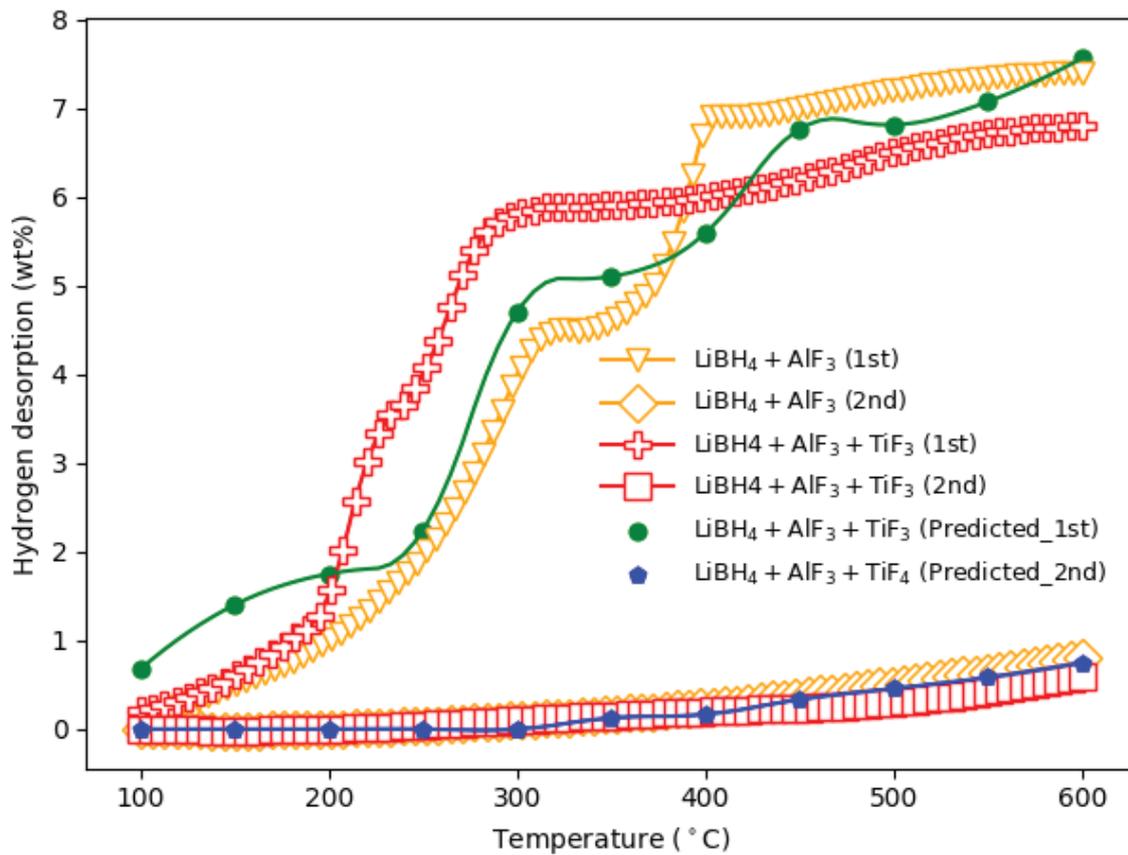


Figure 10: Curves of the first and second dehydrogenation for the $3\text{LiBH}_4 + \text{AlF}_3$ and $3\text{LiBH}_4 + \text{AlF}_3 + 0.2\text{TiF}_3$ composites, respectively (hydrogenation conditions: 450 °C, 9.2 MPa, 24 h).

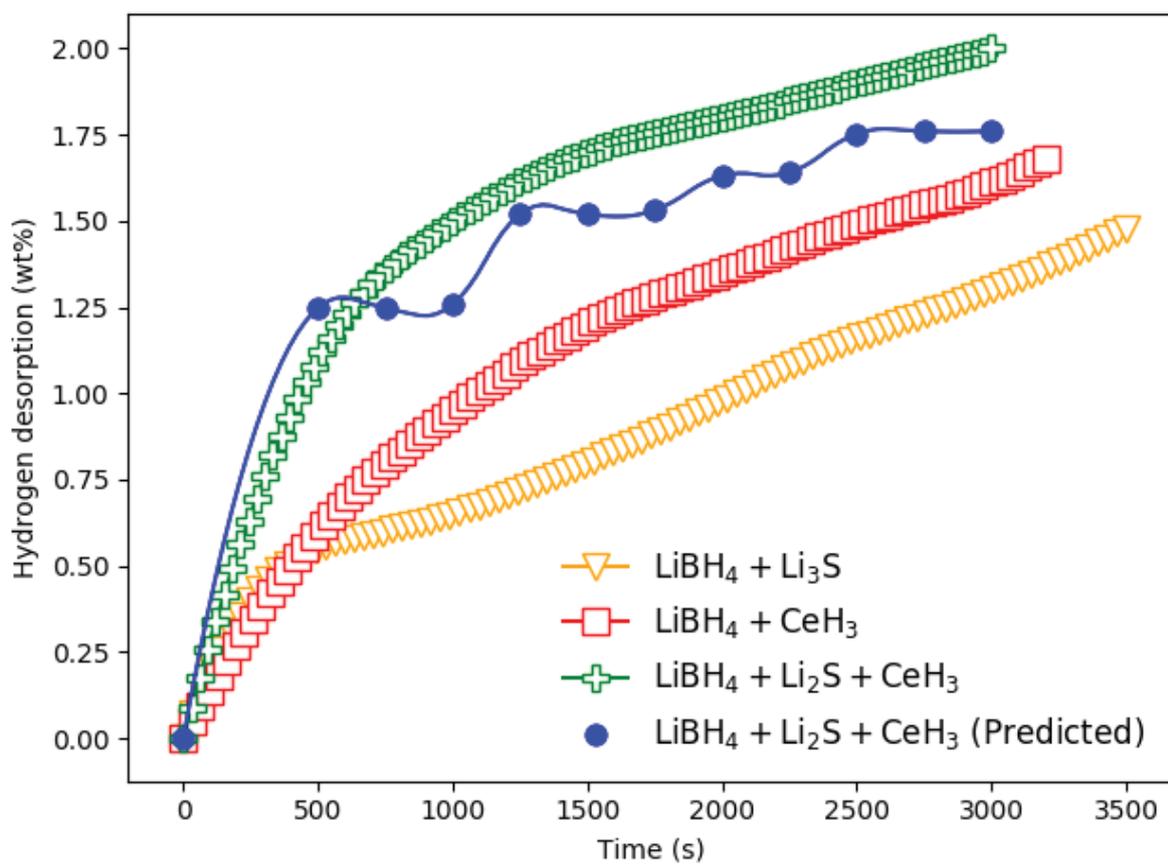


Figure 11: Hydrogen desorption curves of the pristine LiBH_4 , $\text{LiBH}_4 + 20 \text{ wt.}\% \text{Li}_2\text{S}$, $\text{LiBH}_4 + 10 \text{ wt.}\% \text{CeH}_3$ and $\text{LiBH}_4 + 10 \text{ wt.}\% \text{Li}_2\text{S} + 10 \text{ wt.}\% \text{CeH}_3$ mixtures at $350 \text{ }^\circ\text{C}$.

Table 1: Input variables and their ranges.

Variable	Range	Variable	Range
Catalysts	have been listed in Database Construction	Dehydrogenation Atmosphere /bar	0.00001~ 1.5 for Ar 0.0001~ 1.2 for H ₂
Molar Ratio of LiBH ₄ to Catalyst	1:4 ~ 100 : 1	Dehydrogenation Temperature /°C	50 ~ 600
Hand Milling Time/minutes	0 ~ 5	Dehydrogenation Time /hours	0.014~ 300
Ball Milling Time/hours	0 ~ 24	Rehydrogenation Temperature /°C	450~ 600
Ball Milling H ₂ Atmosphere /bar	0 ~ 40	Rehydrogenation H ₂ Pressure /bar	80~ 100
Ball-to-Materials Ratio	0 ~ 180 : 1	Rehydrogenation Time /hours	3 ~ 25
Heating Rate /°C/minute	0 ~ 15	Cycles	1 ~ 5

Table 2: Predictive performance comparison in terms of MSE, RMESE, EV, R^2 , PS and SM correlation. The best two for each metric are marked in bold.

	Linear	Ridge	Lasso	MLP	GBRT	Decision Tree
MSE	3.878±0.235	3.868±0.229	6.158±0.441	3.281±0.633	1.363±0.183	1.826±0.396
RMSE	1.968±0.06	1.966±0.058	2.48±0.09	1.804±0.17	1.165±0.079	1.343±0.15
EV	0.624±0.017	0.625±0.017	0.403±0.029	0.699±0.031	0.868±0.015	0.822±0.039
MAE	1.491±0.039	1.491±0.037	1.97±0.059	1.322±0.167	0.751±0.034	0.657±0.075
R^2	0.622±0.017	0.623±0.017	0.401±0.028	0.681±0.057	0.868±0.015	0.822±0.039
PS	0.791±0.01	0.792±0.01	0.636±0.023	0.838±0.019	0.934±0.009	0.91±0.02
SM	0.839±0.007	0.842±0.008	0.669±0.021	0.863±0.012	0.936±0.006	0.925±0.013
	Ada Decision	Bagging	Extra Trees	RF	Hist	EOE
MSE	1.568±0.312	1.302±0.211	1.24±0.221	1.161±0.171	1.174±0.127	1.144±0.188
RMSE	1.246±0.126	1.138±0.091	1.109±0.098	1.075±0.08	1.082±0.059	1.066±0.087
EV	0.847±0.032	0.873±0.023	0.88±0.021	0.887±0.018	0.886±0.013	0.889±0.019
MAE	0.629±0.049	0.611±0.038	0.535±0.032	0.576±0.028	0.675±0.029	0.601±0.032
R^2	0.847±0.033	0.873±0.023	0.879±0.021	0.887±0.019	0.885±0.013	0.888±0.019
PS	0.922±0.016	0.935±0.012	0.939±0.01	0.942±0.01	0.942±0.007	0.944±0.01
SM	0.936±0.008	0.942±0.007	0.95±0.005	0.948±0.005	0.944±0.006	0.949±0.006

