INTERNATIONAL JOURNAL OF HYDROGEN ENERGY 44 (2019) 7337–7344

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/he

Machine learning based prediction of metal hydrides for hydrogen storage, part I: Prediction of hydrogen weight percent

Alireza Rahnama [a,*], *Guilherme Zepon* [b,**], *Seetharaman Sridhar* [a,c,***]

[a] *ai-garismo, 1 Sandover House, 124 Spa Road, London, SE16 3FD, United Kingdom*
[b] *Department of Materials Engineering, Federal University of São Carlos, Rod. Washington Luis, km 235, CE P: 13565-905, São Carlos-SP, Brazil*
[c] *George S. Ansell Department of Metallurgical and Materials Engineering, Colorado School of Mines, Golden, CO 80401, USA*

ARTICLE INFO

Article history:
Received 16 October 2018
Received in revised form
19 January 2019
Accepted 26 January 2019
Available online 22 February 2019

Keywords:
Machine-learning
Artificial intelligence
Hydrogen storage materials
Metal hydrides

ABSTRACT

The openly available database provided by the US Department of Energy on hydrides for hydrogen storage were analyzed through supervised machine learning to rank features in terms of their importance for determining hydrogen storage capacity referred to as hydrogen weight percent. In this part: we employed four models, namely linear regression, neural network, Bayesian linear regression and boosted decision tree to predict the hydrogen weight percent. For each algorithm, the scored labels were compared to the actual values of hydrogen weight percent. Our investigation showed that boosted decision tree regression performed better than the other algorithms achieving a coefficient of determination of 0.83.

Introduction

With de-carbonization, a hydrogen-based fuel infrastructure becomes attractive and as part of this arises the need to store and transport hydrogen in safe and efficient ways [1,2]. One way is to store hydrogen in solid structures which makes it practical for e.g. transportation. A relatively well studied system is metal hydrides based on various suitable alloy systems [3]. Several sorts of materials, such as hydrides (ionic, covalent or metallic), alanates (e.g. LiAlH4), amides (e.g. Mg(NH2)), N–B compounds (e.g. NH3B3), carbon-based materials and zeolites, are being studied for this application [4]. Among these, metal hydrides remain a promising alternative for solid state hydrogen storage due to their superior safety and higher gravimetric and volumetric storage capacities [5]. In recent years, several alloys and composites have been studied aimed at hydrogen storage applications in the form of metal hydrides. The most important systems and with great potential for technological applications are based on Mg, complex hydrides (mainly based on Mg, Al or Li), base-centered cubic (BCC) solid solutions (such as Ti–Cr–V alloys) and intermetallic

* *Corresponding author.*
** *Corresponding author.*
*** *Corresponding author.* ai-garismo, 1 Sandover House, 124 Spa Road, London, SE16 3FD, United Kingdom.
E-mail addresses: alireza.rahnama@ai-garismo.com (A. Rahnama), zepon@ufscar.br (G. Zepon), sseetharaman@mines.edu, sridhar.seetharaman@ai-garismo.com (S. Sridhar).
https://doi.org/10.1016/j.ijhydene.2019.01.261
0360-3199/

compounds of the AB, A2B, AB2, and AB5 families (for instance, TiFe, Mg2Ni, ZrV2, and LaNi5) [6,7]. The desired hydrogen storage properties are highly dependent on the final application. For example, the gravimetric density is the most important factor for automobile applications, since the hydrogen storage system must suit the size of the vehicle without adding too much weight in order to provide a reasonable driving range. On the other hand, in stationary applications, high volumetric density is preferable, as the weight is not the dominant factor that will affect the efficiency of the system. Another important characteristic of a hydrogen storage system is the temperature of operation, which is the temperature at which the material must work to absorb and desorb hydrogen within a suitable time. Obviously, the lower the operation temperature the higher the efficiency of the storage material. The alloying strategy would be to optimise the H/M (Hydrogen/Metal ratio in the chemical formula for the hydride) ratio to be as high as possible, have an efficient charging and dis-charging capacity and is able to operate as close to ambient temperature and pressure. In addition, cost and raw material availability would become a factor for deployment.

Many metals react with hydrogen to form binary hydrides (MH_n). However, most of the binary hydrides usually do not preset hydrogen storage properties suitable to be used as hydrogen carrier. Some intermetallic compounds can also form hydrides with structural formulae of AB_xH_n, in which element A has a high affinity to hydrogen and is usually a rare earth or an alkaline earth metal and tend to form a stable hydride; and element B has low affinity to hydrogen and is often a transition metal that forms only unstable hydrides. The typical B:A ratios are $x = 0.5$, 1, 2 and 5. There are many examples in literature of intermetallic compounds that form hydrides with a hydrogen to metal ratio of up to two, for instance, Mg_2Ni, Ti_2Ni, TiFe, ZrNi, ZrV_2, $ZrMn_2$, $TiMn_2$, $LaNi_5$, $NaNi_5$ [8,9].

In recent years, informatics approaches are being applied in materials science. Machine learning algorithms has enabled us to utilise past data for rapid predictions and thus it is beginning to change the classical paradigm of materials simulation and computation in which fundamental equations were explicitly solved. Materials informatics methods are extremely useful specifically in cases where reliable data either exist or can be generated for a subset of the critical cases [12]. In the last few years, learning algorithms were employed across a variety of materials including thermoelectrics [13–16], perovskite solids [17–20], carbon-capture materials [21,22], electrocatalysts materials [23,24], oxides and inorganic materials [25–28], the occurrence of interphase precipitation in micro-alloyed steels [29] and light-emitting materials [30]. Previously, machine-learning algorithms were also applied to predict the sorption of C_3H_8, H_2, CH_4 and CO_2 in H_2-selective nanocomposite membranes [31]. In another study, Rezakazemi et al. employed adaptive neuro-fuzzy inference system (ANFIS) for evaluation of H2-selective mixed matrix membranes (MMMs) performance in various operational conditions [32]. Hybrid machine-learning models were also used for the prediction of gas permeation in binary-filler nanocomposite membranes [33]. The possible variations in alloy systems (AB_5, AB_2, etc.) and chemistry variations within each system combined with the multiple operating parameters the system becomes relatively complex for a logical design of an ideal alloy. The alloy properties may be connected, with one-another. Sometimes this is obvious, e.g. temperature and pressure and enthalpy of hydride formation. In other cases, the connections are not so obvious. Due to the large number of permutations and because a reliable data set already exists, the current case is a suitable problem for analysis through data-analytic and machine learning.

Several sorts of materials, such as hydrides (ionic, covalent or metallic), alanates (e.g. $LiAlH_4$), amides (e.g. $Mg(NH_2)$), N–B compounds (e.g. NH_3B_3), carbon-based materials and zeolites, are being studied for this application [4]. Among these, metal hydrides remain a promising alternative for solid state hydrogen storage due to their superior safety and higher gravimetric and volumetric storage capacities [5]. In recent years, several alloys and composites have been studied aiming hydrogen applications in the form of metal hydrides. The most important systems and with great potential for technological applications are based on Mg, complex hydrides (mainly based on Mg, Al or Li), base-centered cubic (BCC) solid solutions (such as Ti–Cr–V alloys) and intermetallic compounds of the AB, A_2B, AB_2, and AB_5 families (for instance, TiFe, Mg_2Ni, ZrV_2, and $LaNi_5$) [6,7]. The desired hydrogen storage properties are highly dependent on the final application. For example, the gravimetric density is the most important factor for automobile applications, since the hydrogen storage system must suit the size of the vehicle without adding too much weight in order to provide a reasonable driving range. On the other hand, in stationary applications, high volumetric density is preferable, as the weight is not the dominant factor that will affect the efficiency of the system. Another important characteristic of a hydrogen storage system is the temperature of operation, which is the temperature at which the material must work to absorb and desorb hydrogen in a suitable time. Obviously, the lower is the operation temperature the higher is the efficiency of the storage material.

We divided our research into two part: on the first part (presented work) we developed regression algorithms to predict the hydrogen weight percent; in the second part, we used classification algorithms to predict the material class based on the properties. In the present study, we firstly identified the correlations between hydrogen weight percent, temperature, heat of formation, material class, composition formulas and pressure. Then, four different regression models, i.e. linear regression, neural network regression, Bayesian linear regression and boosted decision tree regression, were used to predict the hydrogen weight percent as a function of the rest of features we had in the dataset. The tools that we employed in this study were Microsoft Power BI for statistical analysis, Python (Scikit-learn, Matplotlib, Seaborn) for performing correlation analysis and statistical analysis, and Microsoft Azure Machine Learning Studio for regression models and their subsequent performance evaluation.

Dataset

In this paper we utilise the Hydrogen Storage Materials Database, which is an openly available database that can be

accessed through http://hydrogenmaterialssearch.govtools.us/. The creation of this database was funded by the U.S. Department of Energy as a contribution to the International Energy Agency Hydrogen Implementing Agreement (http://ieahydrogen.org/) [10].

This database was created by the Fuel Cell Technologies Office of the US Department of Energy as an initiative to make it easier for users to find up-to-date information on hydrogen and fuel cells. The Hydrogen Storage Materials database is one of three databases created by this organization, in addition to Hydrogen and Fuel Cells Codes and Standards Matrix Database and Hydrogen Safety Bibliographic Database. To the best of our knowledge, the presented work is the very first to use the Hydrogen Storage Materials database to predict hydrogen storage capacities. However, using machine learning approaches to predict materials properties is a research field that has been increasing in the last years [34]. Thermodynamic stability of perovskite, glass transition of temperatures of inorganic glasses, glass forming ability of metallic alloys, among others, are examples of materials properties studied by machining learning approaches [35–38].

This database contains information of hydride materials properties from more than 1600 references. It is important to point out that this database does not contain any information of binary hydrides (MH_n), whose hydrogen properties are well reported in different literature. The Hydride Database classifies the hydrides into nine classes: Mg alloys, Complex hydrides, solid solution alloys, A_2B, intermetallic compounds, AB intermetallic compounds, AB_2 intermetallic compounds, AB_5 intermetallic compounds, and Misc. Intermetallic compounds (MIC), which includes intermetallic compounds that are no classified in the other classes, such as Ce_7Ni_3, $CeMg_{12}$, etc. The total number of reported hydrides is 2722. We removed 4 samples because of missing values that could not be used for the analysis.

The hydrogen storage properties tabulated in the database (used in this work) are: hydrogen storage capacity (in wt.%); enthalpy of hydriding (ΔH) or heat of formation of the hydride; and one selected mid-plateau Pressure-Temperature combination. The latter one is reported as Pressure at 25 °C and Temperature at 1 atm. The Pressure and Temperature entries usually represent extrapolations or interpolations of the Van't Hoff equation. It is worth noting that for very stable hydrides, one may obtain immeasurably low pressure at 25 °C and that values listed should be seen merely as relative indications of room temperature stability. Regarding the hydrogen capacity, it should be mentioned that the tabulated values is given as the fully hydrogenated value, that is, the highest hydrogen concentration measured in the hydride phase limit. It does not necessarily represent the reversible capacity for engineering purposes [11].

Machine learning methods

The dataset was exported to Microsoft Azure Machine Learning Studio as a comma-separated values (csv) file. The *Select Column in Dataset* module was used to manipulate and transform the features into appropriate formats. The *Split Data* module was used to define the training/testing size (0.7/0.3). We used four regression models that are described below in details. After each learning procedure and hyperparameter tuning or training step, a *Score Model* module and *Evaluate Model* module were used to enable us to evaluate the performance of each learning algorithm as well as to make comparisons between the four learning models.

Boosted decision tree regression is a supervised machine learning and is used to create an ensemble of regression trees using boosting [39]. The term *boosting* indicates that each tree depends on prior trees. The learning procedure is to fit the residual of the trees that preceded it. Boosting in a decision tree ensemble, therefore, implies that the learning algorithm tends to enhance accuracy with a small degree of risk of less coverage. Boosting decision tree regression in Azure Machine Learning Studio utilises MART gradient boosting algorithm. We used *Parameter Range* mode in *Create trainer mode* because we aimed to run a parameter sweep. We used *Tune Model Hyperparameters* module that iterates over all the possible combination of parameters to identify the hyperparameters (maximum number of leaves per tree, maximum number of terminal nodes, learning rate which determines how fast or slow the learning algorithm converges to the optimal solution and number of trees) that leads to the optimal results.

The Bayesian linear regression utilises linear regression complemented by additional information in the form of a prior probability distribution [40]. The response from a normal distribution was considered to be $y \sim N(\beta^T X, \sigma^2 I)$. A likelihood function is added to the prior distribution to generate estimates for the parameters, e.g. $P(\beta|y,X) = \frac{P(y|\beta,X)*P(\beta|X)}{P(y|X)}$. Thus, Bayesian linear regression differs from frequentist approach which is represented by standard least-square linear regression with a pre-assumption of data containing sufficient measurements to generate a meaningful model. We used a L2 regularisation with a value of 1 to avoid overfitting. We did not connect the learning algorithm to Tune Model Hyperparameters module as this specific algorithm does not have any parameter that needs tuning.

While neural networks are mainly known for applications in deep learning, they can be easily used to regression problems. Neural network regression is another supervised learning algorithm which requires a label column which must be numerical [41]. Since our label column (hydrogen weight percent) was numerical and it was anticipated that the problem to have a non-linear nature, we used neural network regression. The output of a single neuron has a form of $g(\Sigma_j w_{ji} x_j)$ where w_{ji} are the weights and x_j are the input. A continuous activation function, e.g. sigmoid function of a form of $1/(1+e^{-x})$ was employed. We used *Parameter Range* in the *Create trainer mode* and subsequently employed Tune Model Hyperparameters module to iterate over the possible combinations of parameters to achieve the optimal configuration. For hidden layer specification, we selected a fully-connected case to create a model that has exactly one hidden layer and the output layer is fully connected to the hidden layer and the hidden layer is fully connected to the input layer. Number of nodes in the hidden layer and learning rate were used as the hyperparameters. The final number of hidden

nodes were 100. The optimised value for learning rate was computed to be 0.1. Min-max normalizer was selected to linearly rescale each feature to the [0,1] interval.

We also used the common statistical method [42], linear regression, to predict our targeted numerical value: hydrogen weight percent. A basis function with a form of $g(\vec{w}, \vec{x}) = w_0 + \Sigma_{i=1}^{p} + \Sigma_{i=1}^{p}\Sigma_{k=1}^{p} w_{ik}x_i x_k + \Sigma_{i=1}^{p}\Sigma_{k=1}^{p}\Sigma_{l=1}^{p} w_{ikl}x_i x_k x_l + \ldots$ was used. An ordinary least square was selected as solution method. A 0.001 value was chosen for the L2 regularisation weight. We also used Tune Model Hyperparameters module which in case of linear regression simply trains the learning algorithm.

Results & discussion

Fig. 1a shows the location of samples in a 3D temperature-pressure-hydrogen weight percent space. Most sampled are aggregated in one corner of plot between temperature range 0–200 (°C) pressure 0–25 atm and hydrogen weight percent 0–2 wt%. The figure helped us to identified the outliers, i.e. the samples that are located outside of the clusters. For instance, the sample with hydrogen weight percent around 6% with an operating temperature around 0 °C is an outlier. Outlier are important to identify because they affect the bias (intercept) of the regression models. In the current analysis we classified the outlying points as being located within a normal distribution of residuals or not. If points were found outside the normal distribution of residuals they were discarded. Fig. 1b shows the kernel pairplots for the numerical features. The shadowed areas show the locations with highest probability for any new unseen data. This graph is created based on all the samples within our dataset and it can be biased by the fact that the number of samples in class AB_2 was significantly higher than the others. However, the significance of the kernel pairplots is that it enables us to predict the approximate location of any unseen data in any of these 2-dimensional plane namely: hydrogen weight percent versus heat of formation, temperature and pressure; temperature versus heat of formation and pressure and finally heat of formation versus pressure. The diagonal graphs show the distribution of the corresponding feature for different material class. We also investigated the contourised areas with the highest density of samples. Fig. 2 shows the distribution of properties within each material class, and this indicates what classes offer flexibility in terms of ranges. The highest and lowest values, for each case is provided in Table 1.

Table .2 shows the ranking of features for predicting the value of hydrogen weight percent. For determining the feature importance we used permutation feature importance for which we had an out-of-bag dataset and we randomly permuted the value of the feature j and computed the permuted error ($error_{j,permuted}$) and compared it with overall error ($error_{overall}$). The features were, then, ranked according to the increase in the overall error when they were permuted. As is shown in this table, materials class has the highest rank in determining the value of hydrogen weight percent. Temperature and heat of formation were also identified to be important features with score values of 0.32 and 0.15, respectively. Composition formula was identified to have an insignificant effect on predicting the value of hydrogen weight

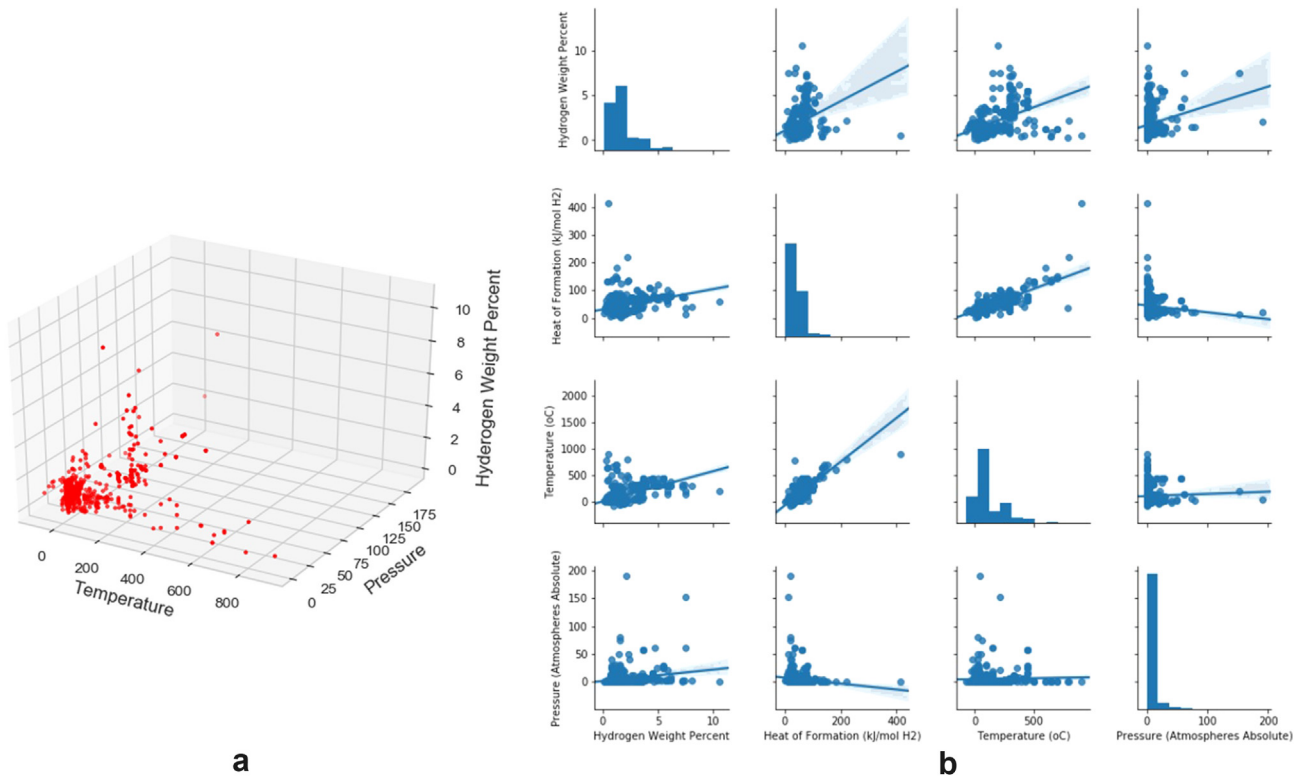

Fig. 1 – (a) 3D scatter plot of data in temperature-pressure-hydrogen weight percent space, (b) Kernel density estimation of the numerical features.

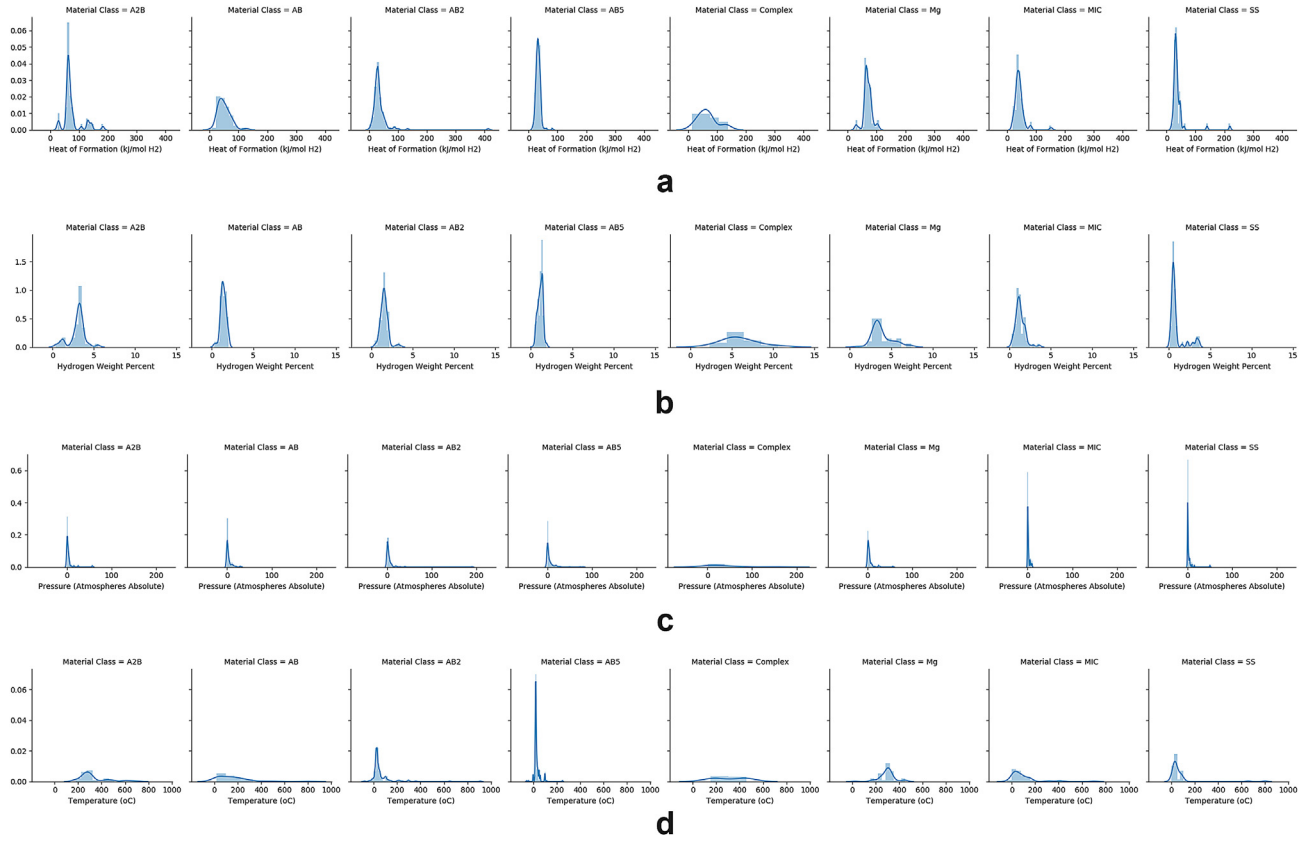

Fig. 2 – The figure shows the distribution of the numerical features for each materials class: (a) Heat of formation, (b) hydrogen weight percent, (c) and (d) temperature.

Table 1 – Feature importance for predicting hydrogen weight percent.

Materials class	Hydrogen weight percent (wt.%)	Heat of formation (kJ/mol H2)	Temperature (°C)	Pressure (atm)
A_2B	0.4–5.5	26.6–183.0	180–700	0.003–57
AB	0.3–2	7.6–126	21–790	0.001–30
AB_2	0.4–3.6	1.03–413	−78–910	0.0001–190
AB_5	0.3–1.9	40.2–82	−55–250	0.01–80
Complex	2.3–10.6	12.8–137	150–458	1–153
MIC	0.3–3.6	19–152	−40–700	0.001–10
Mg	0.9–7.4	26.6–108	25–450	0.004–57
SS	0.1–3.7	17–219	0–800	0.002–50

Table 2 – Feature importance for predicting hydrogen weight percent.

Feature	Materials class	Temperature	Heat of formation	Pressure	Composition formula
Score	0.487527	0.323918	0.15273	0.100549	0.004816

percent. This means that, in other words, in order to predict the value of hydrogen weight percent, it suffices to know the material class, temperature, heat of formation and to some extent pressure for each sample without knowing the corresponding compositional formula.

Fig. 3a is the violin plots of hydrogen weight percent versus material classes. This figure shows that in general complex material class can obtain higher hydrogen weight percent. As shown in this figure the largest variance of hydrogen weight percent belongs to complex material class ranging from 2.3 to 10.6 wt%. It can, also, be seen from this figure that the complex material class has the highest maximum of hydrogen weight percent among all material classes. Mg and A_2B can also reach relatively high hydrogen weight percent after complex material class, while other material classes have a more or less similar range of hydrogen weight percent. Fig. 3b is the bivariate kernel density estimation of hydrogen weight percent versus temperature. As is shown in Fig. 3b, higher hydrogen weight percent, however, comes at a sacrifice of higher temperature, e.g. for a hydrogen weight percent around

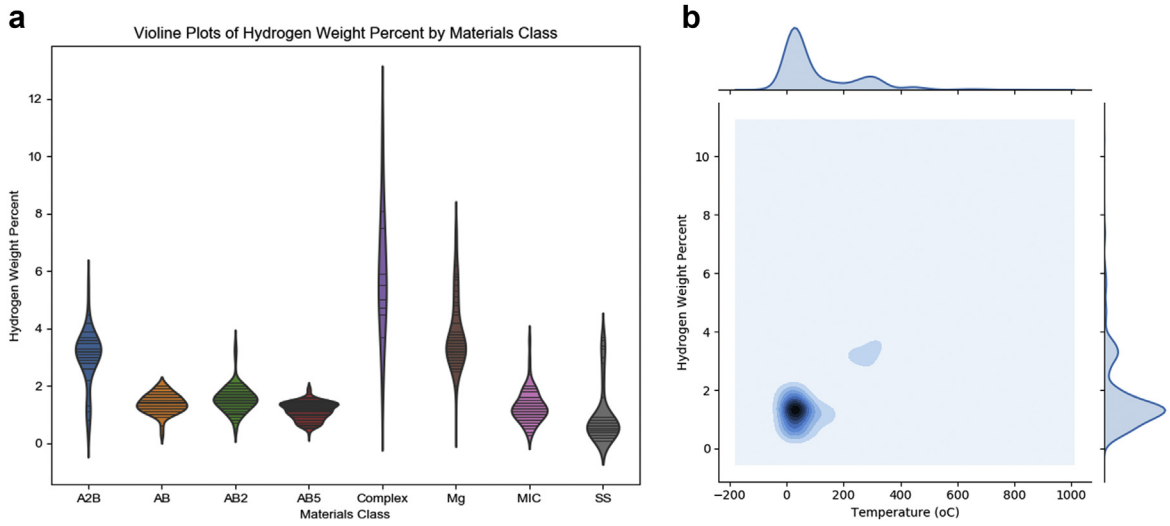

Fig. 3 – (a) Violin plot showing a comparison of hydrogen weight percent for different materials classes, (b) Bivariate kernel density estimation showing the higher hydrogen weight percent comes at a sacrifice of higher temperature.

4 the temperature is almost double when compared to corresponding temperature for hydrogen weight percent with a value of around 2.

The response for the regression models was considered to be the value of hydrogen weight percent. Fig. 4 shows the scatter plots of scored labels versus the actual values of hydrogen weight percent as well as the histogram, cumulative distribution function (CDF) (green) and probability distribution function (PDF) (purple) of values of hydrogen weight percent and scored labels. Among these four models, boosted decision tree regression model performed better than others achieving a coefficient of determination of 0.83. This value for the rest of models was recorded to be 0.56 for Bayesian linear regression, 0.60 for neural network regression and 0.50 for linear regression. The scatter plots also show that scored labels by boosted decision tree regression follows almost the bisector of 2D area. This was not the case for other regression models employed in this study.

The statistical comparison between the actual hydrogen weight percent and the scored labels by boosted decision tree regression revealed that almost all the characteristics of the data were captured by the model. For example, the mean, median, minimum, maximum and standard deviation of value for actual hydrogen weight percent were 1.743, 1.4, 0.1, 8.1 and 1.1193 while those for the scored value were 1.6953, 1.3756, 0.0262, 7.3123 and 1.035, respectively. In addition, the

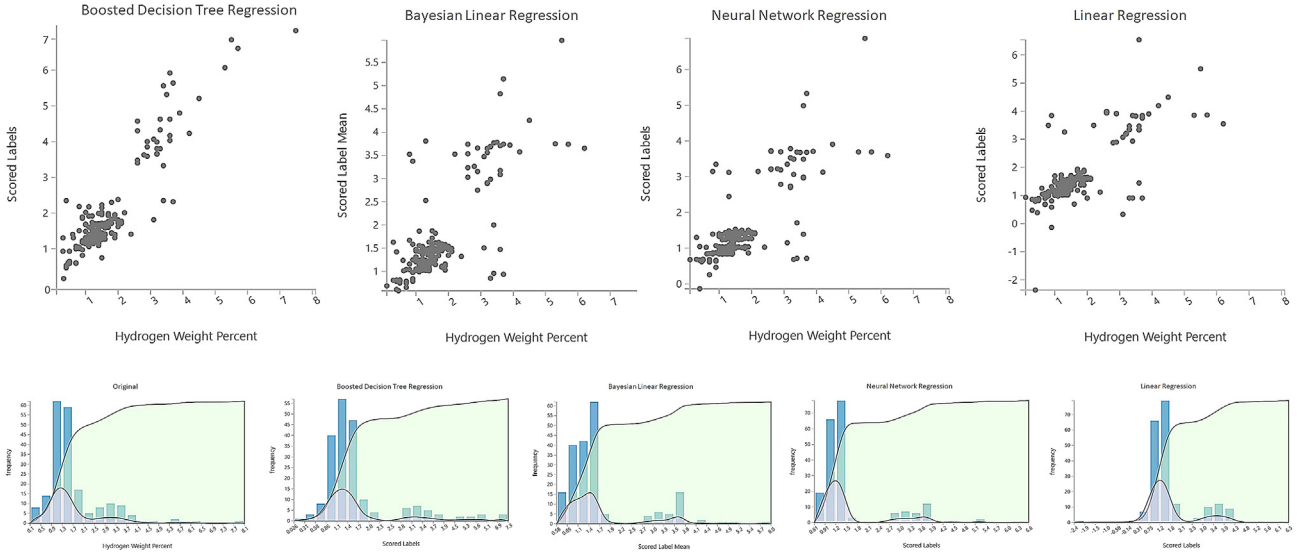

Fig. 4 – The comparison of scored labels versus actual values of hydrogen weight percent for different regression models: boosted decision tree regression, Bayesian linear regression, neural network regression and linear regression; and the corresponding cumulative density function (green) and probability density function (purple) for each regression model compared to the histogram of the actual values of hydrogen weight percent. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3 – Comparison between the statistics of predicted values using four regression models with that of the actual value of hydrogen weight percent.

Statistics	Boosted decision tree regression	Baysian linear regression	Neural network regression	Linear regression	Actual
Mean	1.6953	1.6886	1.556	1.6965	1.743
Median	1.3756	1.4112	1.2685	1.4208	1.4
Min	0.0262	0.5815	−0.1388	−2.3622	0.1
Max	7.6123	5.977	6.8715	6.5398	8.1
Standard Deviation	1.115	0.9864	1.0578	1.0746	1.1193

Table 4 – Evaluation of the performance of four regression models used to predict the value of hydrogen weight percent.

Metric	Boosted decision tree regression	Baysian linear regression	Neural network regression	Linear regression
Mean Absolute Error	0.00303986	0.0442909	0.00482648	0.0417627
Root Mean Squared Error	0.0117862	0.092928	0.0682209	0.0787835
Relative Absolute Error	0.217859	0.571785	0.623087	0.539147
Coefficient of Determination	0.83	0.569095	0.608113	0.502114

relative squared error for boosted decision tree regression was computed to be 0.19 and less than the half of that for other models: 0.43, 0.44 and 0.49 for Bayesian linear regression, neural network regression and linear regression, respectively (Tables 3 and 4).

In addition the CMD and PDF for boosted decision tree closely followed those computed for the actual values of hydrogen weight percent. The PDF for the scored labels using boosted decision tree regression captured the mean of the first distribution (around 1.3) in addition to the subsequent bump with a mean value of 3.3. Among these four regression model, linear regression had the greatest errors and its computed CDF and PDF were well off from the actual functions of the hydrogen weight percent values. This revealed the non-linear nature of this dataset and the non-linear relationship between features - e.g. temperature, heat of formation and pressure - and hydrogen weight percent. In addition, it is well-known that neural network regressors can easily handle high non-linearity. However, the neural network regression model did not perform well for this particular dataset and this could have been due to the model being trapped by local minima instead of progressing towards identifying the global minima.

Conclusion

In summary, we studied the openly available dataset for metal hydrides using machine learning. The results showed, correlation between different features:

- Complex hydrides followed by Mg alloys showed the potential for highest hydrogen weight percent.
- In order to obtain higher hydrogen weight percent, higher temperatures for 1atm of equilibrium pressure are required.
- Among the regressors, boosted decision tree regression performed better with the lowest values of errors and the highest coefficient of determination.
- The statistical analysis of the outcome of regression analyses also confirmed that boosted decision tree regression captured the statistical characteristics of the dataset. The superior performance of boosted decision tree regression and poor performance of linear regression are indicative of the highly non-linear relationship between various variables including temperature, heat of formation, pressure and composition formula with hydrogen weight percent.
- We conducted feature importance analysis and it was found that materials class, temperature, heat of formation highly affect the value of the hydrogen weight percent while composition formula was found to be an insignificant variable for predicting the value of hydrogen weight percent.

Data availability

The dataset is available at [hydrogenmaterialssearch.govtools.us/].

Author contribution

A.R. and S.S. wrote the machine learning algorithms and produced figures. A.R. and S.S. discussed the machine learning results. A.R., S.S. and G.Z. discussed materials case study. All authors wrote and commented on the manuscript and figures.

Acknowledgement

This work was supported by the Serrapilheira Institute (grant number Serra-1709-17362).

REFERENCES

[1] Sinigaglia T, Lewiski F, Santos Martins ME, Mairesse Siluk JC. Production, storage, fuel stations of hydrogen and its utilization in automotive applications-a review. Int J Hydrog

Energy 2017;42:2459724611. https://doi.org/10.1016/j.ijhydene.2017.08.063.
[2] Gurz M, Baltacioglu E, Hames Y, Kaya K. The meeting of hydrogen and automotive: a review. Int J Hydrog Energy 2017;42:24597–611. https://doi.org/10.1016/j.ijhydene.2017.02.124.
[3] Zttel A. Materials for hydrogen storage. Mater Today 2003;6:2433. https://doi.org/10.1016/S1369-7021(03)00922-2.
[4] Chanchetti LF, Oviedo Diaz SM, Milanez DH, Leiva DR, de Faria LIL, Ishikawa TT. Technological forecasting of hydrogen storage materials using patent indicators. Int J Hydrog Energy 2016;41:1830118310. https://doi.org/10.1016/j.ijhydene.2016.08.137.
[5] Libowitz GG. Metallic hydrides; fundamental properties and applications. J Phys Chem Solid 1994;55:14611470. https://doi.org/10.1016/0022-3697(94)90571-1.
[6] Dantzer P. Metal-Hydride technology: a critical review. In: Wipf H, editor. Hydrog. Met. III prop. Appl. Berlin, Heidelberg: Springer Berlin Heidelberg; 1997. p. 279340. https://doi.org/10.1007/BFb0103405.
[7] Chen P, Zhu M. Recent progress in hydrogen storage. Mater Today 2008;11:3643. https://doi.org/10.1016/S1369-7021(08)70251-7.
[8] Westlake DG, Less J. Hydrides of intermetallic compounds: a review of stabilities, stoichiometries and preferred hydrogen sites. Common Mark 1983:1–20. https://doi.org/10.1016/0022-5088(83)90091-7.
[9] Oesterreicher H. Hydrides of intermetallic compounds. Appl Phys 1981:169–86. https://doi.org/10.1007/BF00899753.
[10] De Valladares MR, Jensen JK. The international energy agency hydrogen implementing agreement (IEA HIA): a global perspective on progress and politics in R,D&D cooperation for WHTC2011 conference. In: 4th world hydrog. Technol. Conv. 2011, Glasgow, U.K; 2011. p. 16.
[11] Available from: http://hydrogenmaterialssearch.govtools.us.
[12] Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A, Kim C. Machine learning in materials informatics: recent applications and prospects. Npj Comput Mater 2017;3, 54.
[13] Sparks TD, Gaultois MW, Oliynyk A, Brgoch J. Data mining our way to the next generation of thermoelectrics. Scripta Mater 2016;111:1015.
[14] Yan J, Gorai P, Ortiz B, Miller S, Barnett SA, Mason T, Stevanoviab V, Toberer ES. Material descriptors for predicting thermoelectric performance. Sci 2015;8:983994.
[15] Seshadri R, Sparks TD. Perspective: interactive material property databases through aggregation of literature data. Apl Mater 2016;4:053206.
[16] Oliynyk AO, et al. High-Throughput machine-learning-driven synthesis of full-heusler compounds. Chem Mater 2016;28:73247331.
[17] Pilania G, Balachandran PV, Gubernatis JE, Lookman T. Classification of ABO3 perovskite solids: a machine learning study. Acta Crystallogr B Struct Sci Cryst Eng Mater 2015;71:507513.
[18] Pilania G, Balachandran PV, Kim C, Lookman T. Finding new perovskite halides via machine learning. Front Mater Sci 2016;3:17.
[19] Balachandran PV, Broderick SR, Rajan K. Identifying the 'inorganic gene' for high-temperature piezoelectric perovskites through statistical learning. Proc Roy Soc Lond Math Phys Sci 2011;467:22712290.
[20] Pilania G, et al. Machine learning bandgaps of double perovskites. Sci Rep 2016;6:19375.
[21] Wilmer CE, et al. Large-scale screening of hypothetical metal-organic frameworks. Nat Chem 2011;4:8389.
[22] Lin L-C, et al. Silico screening of carbon-capture materials. Nat Mater 2012;11:633641.
[23] Greeley J, Jaramillo TF, Bonde J, Chorkendorff IB, Nrskov JK. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. Nat Mater 2006;5:909913.
[24] Hong WT, Welsch RE, Shao-Horn Y. Descriptors of oxygen-evolution activity for oxides: a statistical evaluation. J Phys Chem C 2016;120:7886.
[25] Kim E, Huang K, Tomala A, Matthews S, Strubell E, Saunders A, McCallum A, Olivetti E. Machine-learned and codified synthesis parameters of oxide materials. Data Sci 2017;4.
[26] Sumpter BG, Vasudevan RK, Potok T, Kalinin SV. A bridge for accelerating materials by design. Npj Comput Mater 2015;1:15008.
[27] Kalinin SV, GSumpter B, Archibald RK. Big-deep-smart data in imaging for guiding materials design. Nat Mater 2015;14:973980.
[28] Kim E, et al. Virtual screening of inorganic materials synthesis parameters with deep learning. Npj Comput Mater 2017;3, 53.
[29] Rahnama A, Clark S, Sridhar S. Machine learning for predicting occurrence of interphase precipitation in HSLA steels. Comput Mater Sci 2018;154:169–77.
[30] Gmez-Bombarelli R, et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. Nat Mater 2016;15:11201127.
[31] Dashti A, Harami HR, Rezakazemi M. Accurate prediction of solubility of gases within H2-selective nanocomposite membranes using committee machine intelligent system. Int J Hydrog Energy 2018;43:6614–24.
[32] Rezakazemi M, Dashti B, Asghari M, Shirazian S. -selective mixed matrix membranes modeling using ANFIS, PSO-ANFIS, GA-ANFISInt. Int J Hydrog Energy 2017;42:15211–25.
[33] Rezakazemi M, Azarafza A, Dashti A, Shirazian S. Development of hybrid models for prediction of gas permeation through FS/POSS/PDMS nanocomposite membranes. Int J Hydrog Energy 2018;43:17283–94.
[34] Liu Y, Zhao T, Ju W, Shi S, Shi S, Shi S. Materials discovery and design using machine learning. J Mater 2017;3:159–77. https://doi.org/10.1016/j.jmat.2017.08.002.
[35] Ward L, OKeeffe SC, Stevick J, Jelbert GR, Aykol M, Wolverton C. A machine learning approach for engineering bulk metallic glass alloys. Acta Mater 2018;159:102–11. https://doi.org/10.1016/j.actamat.2018.08.002.
[36] Li W, Jacobs R, Morgan D. Predicting the thermodynamic stability of perovskite oxides using machine learning models. Comput Mater Sci 2018;150:454–63. https://doi.org/10.1016/j.commatsci.2018.04.033.
[37] Cassar DR, de Carvalho ACPLF, Zanotto ED. Predicting glass transition temperatures using neural networks. Acta Mater 2018;159:249–56. https://doi.org/10.1016/j.actamat.2018.08.022.
[38] Meredig B, Agrawal A, Kirklin S, Saal JE, Doak JW, Thompson A, Zhang K, Choudhary A, Wolverton C. Combinatorial screening for new materials in unconstrained composition space with machine learning. Phys Rev B Condens Matter 2014;89:1–7. https://doi.org/10.1103/PhysRevB.89.094104.
[39] Available from: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/boosted-decision-tree-regression.
[40] Available from: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/bayesian-linear-regression.
[41] Available from: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/neural-network-regression.
[42] Available from: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/linear-regression.