

Prédiction du stockage de l'hydrogène dans les hydrures métalliques et les hydrures complexes : Une approche d'apprentissage supervisé

HAMMALE Mourad¹ Ikram Firdaous¹ Nouhaila Gargouz¹ Mohamed Benkirane¹
Nawfal Benhamdane¹

Encadrant académique: Hamza Bouhani

¹ Ecole Centrale de Casablanca, Bouskoura, Morocco

Stockage de l'hydrogène; Hydrure Métallique; Chimisorption; Modèle d'apprentissage automatique(ML); Pourcentage de l'hydrogène; Augmentation des données; XGBOOST;

Abstract

This paper investigates the prediction of the hydrogen storage capacity in hydrides using advanced machine learning (ML) models. The most accurate predictions of new materials gravimetric capacity was the XGBoost model. This study achieves a high accuracy in prediction the gravimetric capacity of novel materials using XGBoost model, by enriching basic data with elemental properties. The results reveal correlations between lightweight materials and storage capacity, emphasizing their role in optimizing materials for sustainable energy usage. This work has the purpose to show the potential of combining ML and enriched datasets to accelerate hydrogen storage researchs.

1 Introduction

Dans un monde où l'énergie est le moteur d'une civilisation en perpétuelle évolution, la demande ne cesse d'augmenter. La réponse est bien souvent les énergies à base de combustibles fossiles. Or, ils commencent à se faire de plus en plus rares, et leurs impacts environnementaux tels que la pollution et le changement climatique suscitent de vives préoccupations. Ainsi, la quête d'une transition énergétique qui permettrait de répondre à ces problématiques est essentielle pour atteindre un développement économique et environnemental durable[1].

L'hydrogène répond alors à ces demandes comme étant une alternative énergétique respectueuse de l'environnement tout en offrant une efficacité de conversion énergétique et de production. C'est une ressource abondante qui peut être extraite avec une fraction massique d'environ 11% dans l'eau. Son gaz possède une densité énergétique de 1.42×10^8 joules par kilogramme, soit 2 à 3 fois la valeur calorifique de l'essence. Il est aussi important de noter que l'hydrogène est une option flexible pour le transport de l'énergie sur de longues durées, rendant la production écoresponsable de l'énergie plus accessible.

En vue de ses bénéfices, le Conseil de l'hydrogène prévoit une augmentation de la production de l'hydrogène propre de 800 000 à 38 millions de tonnes par an d'ici 2030. La Chine, les États-Unis, l'Union européenne et le Japon ont déjà commencé à mettre en place des stratégies de développement d'énergie hydrogène adaptées à leurs propres besoins[1].

Ceci dit, l'énergie hydrogène rencontre toujours des défis, notamment la production efficace à partir de sources d'énergie renouvelables, le stockage à grande échelle et le coût toujours élevé. Parmi ces défis, le plus pressant est celui du stockage sûr et efficace. Ceci permettra de fournir une gestion fluide de l'énergie en assurant sa continuité et sa stabilité. Face à ce défi, le stockage de l'hydrogène à l'état solide se présente comme l'une des méthodes de stockage les plus sûres et avec une haute efficacité énergétique. Il permet de réduire considérablement les risques de fuite ou d'explosion associés au stockage de l'hydrogène à l'état gazeux ou liquide. Cependant, les matériaux de stockage de l'hydrogène à l'état solide se caractérisent par une grande variété de types et de performances, imposant alors des processus expérimentaux complexes et coûteux.

Afin d'accélérer la recherche dans ce domaine encore technologiquement novice mais prometteur, plusieurs méthodes ont été introduites, **Théorie de la fonctionnelle de la densité (DFT) et l'apprentissage automatique (ML)**.

La DFT est un outil puissant qui permet, en analysant des propriétés telles que les modes vibratoires ou les spectres énergétiques des matériaux, de générer des données riches et d'extraire des caractéristiques clés. Ces dernières sont ensuite exploitées grâce à l'apprentissage automatique, qui permet de les analyser et facilite leur interprétation en détectant des corrélations entre les propriétés des matériaux tout en ouvrant la porte à une éventuelle prédiction de données et de structures pas encore calculées. De ce fait, la combinaison de ces deux technologies s'avère être un terrain fertile pour l'accélération d'un domaine prometteur : le stockage de l'hydrogène à l'état solide.

Dans cette synthèse, nous allons nous consacrer à l'étude du stockage de l'hydrogène à l'état solide via la chimisorption. Le but sera d'utiliser des données relatives aux hydrures complexes et aux hydrures métalliques et de les utiliser pour la prédiction de la capacité de stockage de l'hydrogène dans des matériaux non encore exploités à travers **une approche ML**. Cette démarche permet d'identifier des pistes à grand potentiel avant d'engager l'expérimentation. Pour ce faire, nous allons commencer par une étude analytique de l'absorption de l'hydrogène.

1.1 Stockage de l'hydrogène à l'état solide

Le stockage solide de l'hydrogène consiste à conserver l'hydrogène dans d'autres matériaux solides. Cela se fait soit par chimisorption, qui est une liaison forte covalente, ionique ou métallique, soit par physisorption, qui se fait grâce à des liaisons faibles, effets de forces intermoléculaires de Van der Waals. Les caractéristiques des matériaux qu'il faut prendre en considération sont la haute densité gravimétrique et volumique, la rapidité d'absorption et de désorption en conditions ambiantes, le coût, la disponibilité, la résistance aux impuretés de l'hydrogène et la capacité d'être utilisés plusieurs fois[2]. Les matériaux utilisés pour le stockage d'hydrogène peuvent être classés comme suit :

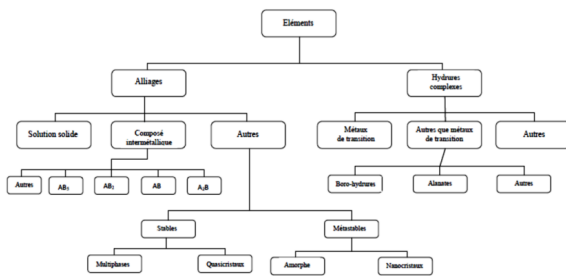


Figure 1. Classification des éléments de stockage solide d'hydrogène [3].

Dans cette synthèse, nous allons nous consacrer aux matériaux qui permettent un stockage solide de l'hydrogène via la chimisorption.

1.1.1 Les hydrures complexes

Avec une capacité massique de l'ordre de 10%, les hydrures complexes présentent un grand potentiel pour le stockage solide de l'hydrogène.

1.1.1.1 Les alانات

Ils offrent une liaison ionique entre un complexe anionique $[AlH_{n+3}]^{n-}$ et un cation métallique M^n , alcalin ou alcalino-terreux. Ils peuvent également être dopés par un certain matériau de transition (Ti, Fe). Les alانات peuvent absorber et désorber jusqu'à 5,6% en masse d'hydrogène. Cependant, ils nécessitent des températures élevées pour procéder à la désorption, ce qui limite leur utilisation[3].

1.1.1.2 Les amides

Ils se forment à partir des liaisons covalentes entre l'azote et l'hydrogène. Ceci se fait par formation successive des imidures et des amidures, permettant ainsi une bonne absorption de l'hydrogène. En l'occurrence, l'amidure de lithium $LiNH_2$ a une capacité massique théorique de 12%. Cependant, ce genre de stockage est désavantageux en raison de son explosivité en cas de présence d'eau[3].

1.1.1.3 Les borohydrures

Ils se composent grâce à une liaison entre des atomes de bore liés à des atomes d'hydrogène et un cation métallique ou un groupement organique sous la structure $M[BH_4]$. En général, ce type de matériau a une bonne capacité de stockage, comme le cas de $LiBH_4$, le borohydru de lithium avec 18% de capacité de stockage hydrogène en masse. Cependant, leurs températures de stockage et de déstockage sont souvent très élevées pour une utilisation pratique[3].

1.1.2 Hydrures métalliques

Les métaux peuvent former des hydrures à partir de liaisons ioniques, covalentes ou métalliques avec l'atome d'hydrogène. Les atomes d'hydrogène s'intègrent dans les structures cristallines du métal, en occupant soit les sites interstitiels, soit les sites substitutifs. Les composés intermétalliques sont généralement composés d'un élément A qui forme un hydru de stable et un élément B qui forme un hydru de instable. Ceci permet de les classer en familles réversibles et en familles irréversibles.

La réaction des hydrures métalliques réversibles est de la forme suivante :



Les composés métalliques qui interagissent avec l'hydrogène suivent un mécanisme en cinq étapes :

- (1) Adsorption de l'hydrogène gazeux à la surface du matériau.
- (2) Réduction des couches de passivation éventuelles, telles que les oxydes, nitrures ou sulfures, présentes sur la surface.
- (3) Dissociation des molécules d'hydrogène en atomes individuels.
- (4) Absorption (ou chimisorption) des atomes d'hydrogène à la surface du matériau.
- (5) Diffusion des atomes d'hydrogène vers le cur du matériau. Les atomes d'hydrogène migrent progressivement à travers le réseau métallique, leur fréquence de saut étant régie par la température selon la loi de Fick.

Les hydrures métalliques présentent de nombreux avantages pour le stockage de l'hydrogène. En offrant une densité volumique élevée, atteignant jusqu'à 50-60 kg/m³, surpassant celle de l'hydrogène gazeux ou liquide, comme le démontrent des matériaux tels que le MgH_2 (7,6% en masse) ou les alliages $LaNi_5$ et $TiFe$, ils sont l'un des candidats les plus prometteurs. Leur cinétique rapide d'absorption et de désorption à température ambiante ($\sim 25^\circ C$) et à des pres-

sions modérées (~ 10 bars) est exemplifiée par le LaNi_5 , tandis que le TiFe combine une cinétique compétitive et un faible coût[3].

En outre, leur sécurité intrinsèque à l'état solide réduit considérablement les risques de fuite ou d'explosion, ce qui est essentiel pour des applications sensibles comme les véhicules ou les infrastructures résidentielles. Ces matériaux se distinguent également par leur stabilité et leur réversibilité, permettant des cycles répétés sans dégradation notable, comme le montre le LaNi_5 , capable de maintenir ses propriétés sur des centaines de cycles.

La diversité des compositions disponibles, comme AB_5 , AB_2 ou AB , permet d'adapter leurs performances à différents besoins, allant de l'automobile à l'énergie stationnaire. En outre, leur coût modéré, grâce à l'utilisation d'éléments abondants comme le Mg et le Fe, les rend économiquement viables pour une adoption à grande échelle. Enfin, leur robustesse et leur compatibilité avec les applications industrielles en font des matériaux stratégiques pour des systèmes variés, tels que les dispositifs portables, les infrastructures fixes ou les projets liés aux énergies renouvelables.

2 Méthodologie

2.1 Collecte des données

Les données de cette étude sont issues d'une base de données ML_HYDPARK[4] qui a été mise à jour en février 2024. Elle contient des caractéristiques expérimentales détaillées sur les hydrures métalliques, à savoir la composition chimique, les propriétés thermodynamiques notamment l'enthalpie de formation et l'entropie. On a choisi comme variable de sortie la capacité massique de stockage d'hydrogène (wt%), vu que c'est un paramètre essentiel pour prédire les performances des matériaux.

2.2 Analyse initiale et exploration des données

Une analyse exploratoire des données nous a permis d'étudier les relations entre les caractéristiques et la capacité massique (wt%). Parmi celles analysées: l'enthalpie ΔH , l'entropie ΔS , la température et la pression. Une première tentative de prédiction de la variable de sortie a été réalisée à partir des modèles d'apprentissage supervisé. Cependant, les résultats obtenus ont donné des performances faibles, ce qui montre la nécessité d'employer une nouvelle approche permettant d'ajuster les modèles.

2.3 Affinage des données et ingénierie des caractéristiques

Pour améliorer les performances des modèles prédictifs, une étape d'affinage des données a été entreprise. Cette démarche a inclus :

- (1) **Analyse des caractéristiques** : Les caractéristiques présentant une faible corrélation avec wt% ou une pertinence théorique limitée ont été écartées. De nouvelles caractéristiques fortement corrélées avec wt% ont été ajoutées. Ces caractéristiques supplémentaires, validées théoriquement, incluent des descripteurs dérivés des propriétés élémentaires, des paramètres thermodynamiques et des conditions de traitement.
- (2) **Nettoyage des données** : La base de données con-

tient des valeurs manquantes dans certaines de ses caractéristiques. On les a donc rempli par des méthodes de régression linéaire, et ceci en identifiant les corrélations pouvant exister entre les caractéristiques afin d'obtenir des valeurs fiables

- (3) **Normalisation** : Les variables ont ensuite été normalisées afin de les aligner sur des échelles comparables, ce qui améliore leur compatibilité avec les algorithmes d'apprentissage automatique et optimise les performances des modèles prédictifs.

2.4 Réévaluation des modèles prédictifs

Après avoir affiné les données, une nouvelle phase de modélisation a été initiée. Divers algorithmes d'apprentissage supervisé, tels que les forêts aléatoires, CatBoost, XGBoost et les réseaux neuronaux, ont été testés. Chaque modèle a été évalué sur des critères de précision, notamment l'erreur quadratique moyenne (MSE), l'erreur absolue moyenne (MAE) et le coefficient de détermination (R^2). Ces indicateurs ont permis de comparer rigoureusement les performances et de sélectionner les approches les plus adaptées aux objectifs de l'étude.

3 Résultats et Discussion

3.1 Analyse et description des données

Notre étude a débuté par une analyse des différentes caractéristiques décrivant les propriétés physiques, chimiques et thermodynamiques des matériaux étudiés pour le stockage de l'hydrogène. Ceci inclue des paramètres comme la classification des matériaux (*Material_Class*), le regroupant des catégories comme A_2B , AB ou Mg , définies par des structures chimiques spécifiques influençant leur comportement vis-à-vis de l'hydrogène.

En effet, la capacité massique de stockage de l'hydrogène, la quantification de l'énergie de liaisons ou de rupture avec l'hydrogène et la pression d'équilibre à 25 C sont donnés respectivement par les caractéristiques suivantes: *Hydrogen_Weight_Percent*, *Heat_of_Formation_kJperMolH2* et (*Equilibrium_Pressure_25C*).

Nous avons noté que la caractéristique *Equilibrium_Pressure_25C* a une tendance variationnelle forte. Pour cela nous avons opté par une version transformée de cette dernière: *LnEquilibrium_Pressure_25C*.

Cette transformation a compressé l'échelle des valeurs, réduit l'impact des extrêmes et rendu la distribution plus uniforme, facilitant ainsi les analyses statistiques. Thermodynamiquement, cette transformation est justifiée par l'équation de Van't Hoff :

$$\ln(P_{eq}) = -\frac{\Delta H}{RT} + \frac{\Delta S}{R}$$

où $\ln(P_{eq})$ est le logarithme de la pression d'équilibre, ΔH la chaleur de formation, T la température absolue, R la constante universelle des gaz, et ΔS l'entropie. Cette relation montre que *LnEquilibrium_Pressure_25C* est directement liée à ΔH , permettant ainsi de linéariser les relations complexes.

L'analyse a révélé une corrélation négative forte (-0.94) entre *LnEquilibrium_Pressure_25C* et *Heat_of_Forma-*

tion

$kJperMolH_2$, confirmant que des matériaux ayant une chaleur de formation élevée, et donc des liaisons fortes, présentent des pressions d'équilibre faibles.

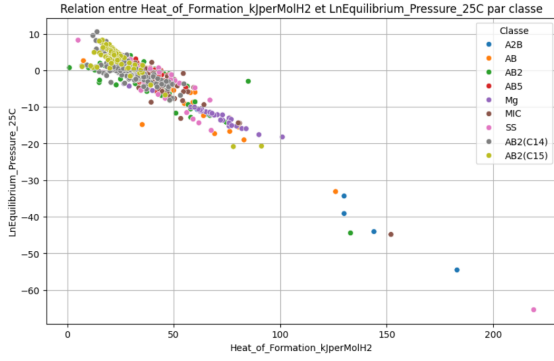


Figure 2. Relation entre $Heat_of_Formation_kJperMolH_2$ et $LnEquilibrium_Pressure_25C$ par classe.

Nous avons choisi d'étudier la corrélation entre les features pour comprendre les relations sous-jacentes entre les propriétés physiques et thermodynamiques des matériaux, afin d'identifier les mécanismes clés qui influencent leur performance et d'orienter leur optimisation pour des applications spécifiques, telles que le stockage d'hydrogène, dans notre cas. Cette approche a ensuite permis d'explorer des corrélations importantes entre les autres variables. Par exemple, nous avons étudié la corrélation entre $Heat_of_Formation_kJperMolH_2$ et $Entropy_of_Formation_JperMolH_2perK$, révélant une relation positive où une augmentation de la chaleur de formation ΔH s'accompagne d'une augmentation de l'entropie ΔS .

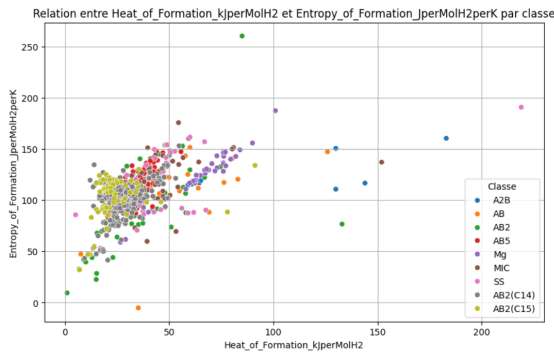


Figure 3. Relation entre $Heat_of_Formation_kJperMolH_2$ et $Entropy_of_Formation_JperMolH_2perK$ par classe.

3.2 Augmentation de Data

Afin de mieux prédire la capacité de stockage de l'hydrogène des matériaux et d'identifier ceux présentant les meilleures performances, nous avons utilisé la bibliothèque Mendeleev. Cette bibliothèque génère automatiquement des propriétés liées à leurs propriétés atomiques. Cette approche nous a permis de créer de nouvelles caractéristiques fournissant une description détaillée des propriétés chimiques, physiques et électroniques des matériaux.

Cette analyse repose sur neuf propriétés principales : numéro atomique, période, électronégativité, affinité élec-

tronique, volume et masse atomiques, chaleur de fusion, énergie d'ionisation et rayon covalent. Ces paramètres permettent de comprendre les mécanismes chimiques et thermodynamiques du stockage de l'hydrogène. Nous avons réalisé des calculs statistiques comprenant des sommes pondérées, des écarts-types pour mesurer la variabilité et des bornes minimales et maximales pour définir les limites des matériaux.

L'importance de ces caractéristiques réside dans leurs significations physiques impactant le stockage de l'hydrogène. A titre d'exemple, le numéro atomique et la période font référence à la taille et la structure atomique des matériaux ce qui influence directement leur capacité à interagir avec l'hydrogène. L'électronégativité et l'affinité électronique sont des indicateurs de la stabilité et de la force des liaisons chimiques, tandis que le volume atomique et la masse atomique influencent respectivement la capacité volumétrique et gravimétrique. La chaleur de fusion et l'énergie d'ionisation donnent des informations sur la stabilité thermique et chimique des matériaux. Le rayon covalent a un impact direct sur la structure cristalline et les espaces disponibles pour accueillir les molécules d'hydrogène.

En intégrant ces nouvelles features, nous avons étudié leurs corrélations avec $Hydrogen_Weight_Percent$, qui mesure la capacité de stockage d'hydrogène. Cette analyse a permis d'identifier les dix caractéristiques les plus importantes dont les sommes pondérées de la période et du numéro atomique, ainsi que les valeurs maximales et minimales de ces propriétés, jouent un rôle central. Voici les résultats obtenus :

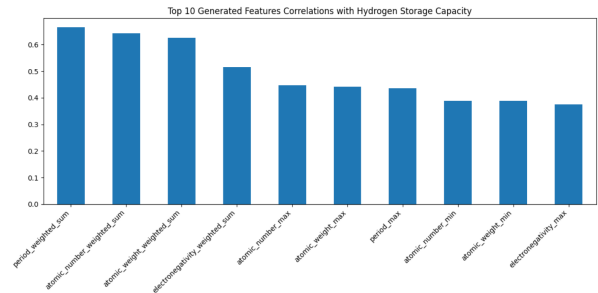


Figure 4. Top 10 des corrélations avec $Hydrogen_Weight_Percent$.

Top 10 Features Générées :

Feature	Importance
period_weighted_sum	0.665297
atomic_number_weighted_sum	0.643065
atomic_weight_weighted_sum	0.624565
electronegativity_weighted_sum	0.515793
atomic_number_max	0.447632
atomic_weight_max	0.442187
period_max	0.434762
atomic_number_min	0.388731
atomic_weight_min	0.388343
electronegativity_max	0.375022

Table 1. Top 10 des features g n r es les plus importantes pour le stockage d'hydrog ne.

Ces r sultats d montrent que les propri t s pond r es, ainsi que les valeurs extr mes des propri t s  l mentaires, sont d terminantes pour la capacit  d'un mat riau   stocker l'hydrog ne. Afin de mieux visualiser les relations entre ces features s lectionn es et *Hydrogen_Weight_Percent*, nous avons construit une matrice de corr lation, pr sent e sous forme de carte thermique. Cette matrice, illustr e dans la figure 5, met en  vidence les corr lations significatives entre les principales variables.

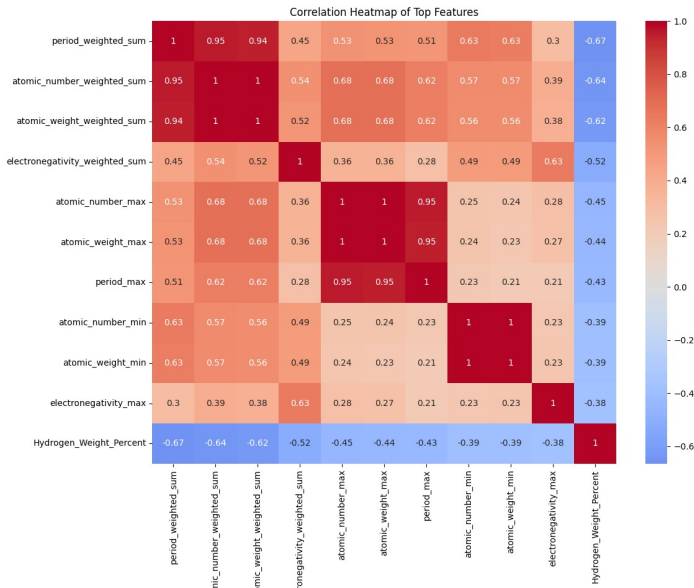


Figure 5. Matrice de corr lation des principales features.

La matrice de corr lation nous d montre des forts liens entre les diverses propri t s des mat riau et leur aptitude au stockage de l'hydrog ne. Nous observons que les  l ments des premi res p riodes du tableau p riodique (c'est- -dire les p riodes 2 et 3), qui sont les plus l gers sont les mieux adapt s pour augmenter la capacit  de stockage. Ceci est d montr  par des corr lations n gatives significatives comme la p riode pond r e (-0.67) ou au num ro atomique pond r  (-0.64). De m me, la relation avec la

masse atomique pond r e (-0,62) montre l'importance des  l ments l gers comme le lithium et le magn sium, qui sont bien utilis s fr quemment dans les technologies de stockage..

Le num ro atomique, la masse atomique et la p riode ont un taux de corr lations qui atteint 0.9 ce qui montre une interd pendance naturelle dans la classification p riodique. En addition   la corr lation avec l lectron gativit  de -0,52, jouant ainsi un r le notable en  quilibrant la stabilit  chimique et la lib ration d'hydrog ne.

Les r sultats confirment alors que les  l ments l gers, majoritairement issus des premi res p riodes, peuvent  tre potentiellement une base solide pour concevoir des mat riau performants en tenant compte de la gravim trie et de la stabilit  chimique.

3.3 Mod les de Pr diction ML utilis s et r sultats obtenues

Dans cette synth se, nous avons  tudi  et compar  plusieurs mod les de machine learning dans le but d'identifier le mod le le plus ad quat pour pr dire avec pr cision la capacit  de stockage d'hydrog ne de chaque mat riau. Les mod les  valu s incluent XGBoost, Random Forest, CatBoost et un r seau de neurones. Pour chacun deux, les performances ont  t  mesur es   l'aide de m triques telles que le score R^2 , le RMSE (Root Mean Squared Error) et le MAE (Mean Absolute Error), sur des ensembles d'entra nement et de test. Notre objectif principal  tait de s lectionner le mod le offrant le meilleur score R^2 , garantissant ainsi une haute pr cision dans les pr dictions. Ce mod le permettra non seulement de pr dire efficacement la capacit  de stockage d'hydrog ne mais aussi de visualiser et d'identifier les mat riau pr sentant une grande performance, afin de mieux comprendre leurs caract ristiques et de guider leur optimisation pour des applications futures. La figure 6 illustre les valeurs pr dites par chaque mod le par rapport aux valeurs r elles, offrant une vue d'ensemble de leur performance respective. XGBoost :

XGBoost :

- Ensemble d'entra nement : $R^2 = 0.926$, MSE = 0.040, RMSE = 0.201, MAE = 0.145.
- Ensemble de test : $R^2 = 0.926$, MSE = 0.040, RMSE = 0.201, MAE = 0.145.

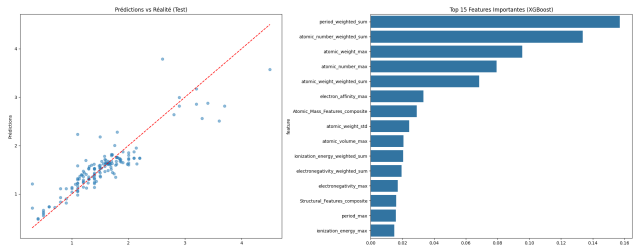


Figure 6. XGBoost : Valeurs pr dites vs Valeurs r elles.

Random Forest :

- Ensemble d'entra nement : $R^2 = 0.928$, RMSE = 0.199, MAE = 0.130.
- Ensemble de test : $R^2 = 0.779$, RMSE = 0.306, MAE = 0.198.

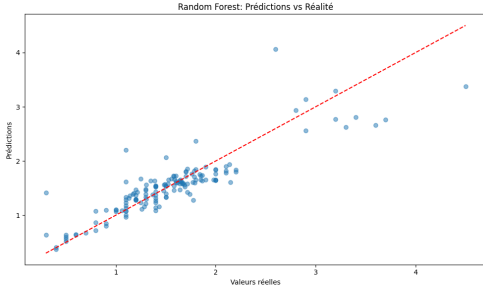


Figure 7. Random Forest : Valeurs prédites vs Valeurs réelles.

CatBoost :

- Ensemble d'entraînement : $R^2 = 0.978$, RMSE = 0.110, MAE = 0.085.
- Ensemble de test : $R^2 = 0.793$, RMSE = 0.295, MAE = 0.190.

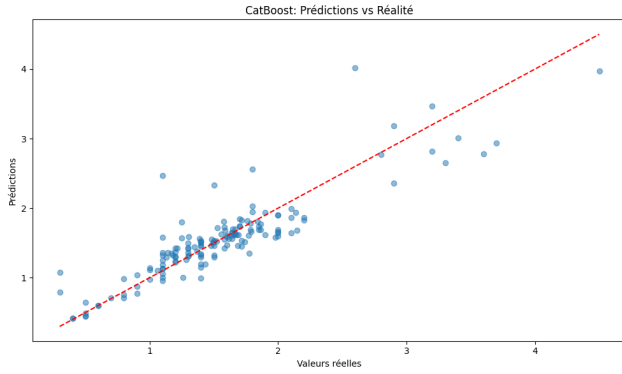


Figure 8. CatBoost : Valeurs prédites vs Valeurs réelles.

Réseau de Neurones :

- Ensemble d'entraînement : $R^2 = 0.587$, RMSE = 0.476, MAE = 0.328.
- Ensemble de test : $R^2 = 0.500$, RMSE = 0.459, MAE = 0.330.

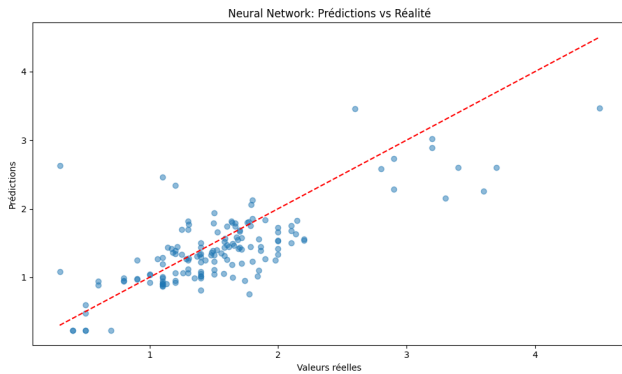


Figure 9. Réseau de Neurones : Valeurs prédites vs Valeurs réelles.

Comparaison des Modèles :

Modèle	R^2 Train	RMSE Test	R^2 Test
XGBoost	0.926	0.201	0.926
Random Forest	0.928	0.306	0.779
CatBoost	0.978	0.295	0.793
Réseau de Neurones	0.587	0.459	0.500

Figure 10. Comparaison des performances des modèles[5].

En analysant les performances, XGBoost s'est révélé être le modèle le plus performant, avec un excellent équilibre entre précision et généralisation, surpassant les autres modèles sur l'ensemble de test. L'identification des matériaux capables de stocker efficacement l'hydrogène constitue une étape clé pour accélérer le développement de nouvelles solutions tout en réduisant les coûts et en optimisant les processus industriels. En ciblant les matériaux les plus prometteurs, cette démarche permet de maximiser la capacité gravimétrique tout en garantissant une bonne stabilité chimique, contribuant ainsi à des applications énergétiques durables. C'est dans cette optique que nous avons exploité notre modèle **XGBoost** pour prédire la capacité de stockage d'hydrogène des matériaux présents dans notre jeu de données. Cette prédiction a été suivie d'une visualisation des résultats afin d'identifier les matériaux présentant les meilleures performances et de mieux comprendre leurs caractéristiques intrinsèques. Ce travail permet non seulement d'orienter les efforts de recherche vers les composés les plus prometteurs, mais également de relever des défis critiques tels que l'optimisation de la densité énergétique, la stabilité thermique et la faisabilité économique des technologies basées sur l'hydrogène. La figure fig:violin_plot_hydrogen, obtenue, présente la distribution des capacités de stockage d'hydrogène (en %) par classe de matériau, mettant en évidence les variations et les tendances associées à chaque classe étudiée :

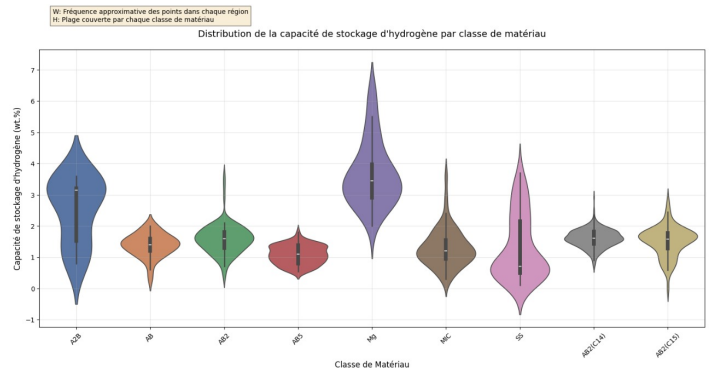


Figure 11. Distribution de la capacité de stockage d'hydrogène (en %) par classe de matériau.

3.4 Discussion des résultats

Pour mieux évaluer notre approche, nous avons comparé nos résultats avec ceux d'une étude similaire,[6], mais qui n'a pas intégré de nouvelles features dans son jeu de données.

Le tableaux ci-dessous résume les performances des modèles évalués dans les deux études. Une comparaison directe

des modèles communs permet de mettre en évidence les avantages de notre méthodologie.

Metric Model	Mean square error	Mean/Relative absolute error	Coefficient of determination (R^2)
Linear	0.08	0.54	0.50
Regression	0.42	0.43	0.64
Polynomial regression	1.24	0.75	0.55
Random forest regression	0.90	0.59	0.67
Decision tree regression	0.19	0.35	0.93
Boosted decision tree	0.01	0.21	0.83
Neural network regression	0.07	0.62	0.54
Multiclass logistic regression	-	-	0.47
Multiclass decision forest	-	-	0.60
Multiclass decision jungle	-	-	0.62
Multiclass neural network	-	-	0.80

Figure 12. Performances des modèles présentés dans l'article étudié

En comparant les modèles communs, il apparaît clairement que notre approche, basée sur l'ajout de nouvelles *features* générées à partir des propriétés élémentaires, a significativement amélioré les performances des modèles.

Par exemple, pour *Random Forest*, notre étude a obtenu un $R^2_{\text{Test}} = 0.779$, bien supérieur au 0.67 obtenu dans l'article. Cette amélioration considérable illustre l'impact de la *Data Augmentation*, qui permet au modèle de mieux capturer les relations complexes entre les propriétés des matériaux (thermodynamiques, atomiques) et leur capacité de stockage d'hydrogène.

De même, les résultats pour le réseau de neurones révèlent une petite différence. Dans l'article, ce modèle a obtenu un $R^2 = 0.54$ avec une MSE de 0.07, alors que dans notre étude, le même modèle affiche un $R^2_{\text{Test}} = 0.500$. Cette divergence peut être expliquée par une différence dans la taille et la qualité des données utilisées, notre modèle étant basé sur des données enrichies mais sans optimisation spécifique pour cette méthode.

Il est évident, en examinant les performances globales, que notre approche offre un grand avantage par rapport aux modèles linéaires présentés dans l'article. Par exemple :

- La régression linéaire ($R^2 = 0.50$),
- La régression polynomiale ($R^2 = 0.55$),

affichent des performances nettement inférieures comparées à nos modèles avancés, tels que le modèle XGBoost ($R^2_{\text{Test}} = 0.926$).

Ces résultats démontrent que l'ajout de nouvelles *features* permet une meilleure capture des relations non linéaires et complexes entre les propriétés des matériaux.

Enfin, l'intégration de nouvelles *features* via la bibliothèque *Mendeleev* a joué un rôle essentiel dans cette amélioration. Les propriétés générées, telles que les sommes pondérées et les valeurs extrêmes des propriétés élémentaires, permettent de mieux comprendre et modéliser les interactions complexes entre les matériaux et leur capacité à stocker l'hydrogène.

3.5 Analyse visuelle des distributions

Maintenant, les courbes de distribution de la capacité de stockage d'hydrogène par classe sont comparées aux résultats obtenus dans cette étude et à ceux présentés dans l'article de référence[6]. Ces visualisations sont importantes pour évaluer l'impact des approches méthodologiques sur la précision de nos prédictions.

La visualisation obtenue représentée à la figure 13, montre un équilibre au niveau de la répartition des capacités de stockage pour chaque classe. Les matériaux type Mg présentent de larges intervalles, ce qui se justifie par leur grande performance au niveau du stockage d'hydrogène. Les classes A_2B et SS affichent aussi des distributions bien distinctes, avec des intervalles interquartiles étroits. Tout cela reflète la capacité de notre approche à détecter les relations complexes entre les propriétés des matériaux et leur capacité de stockage..

En revanche, la courbe issue de l'article montre une dispersion importante des données, avec des distributions peu distinctes entre les classes. Cette dispersion, visible par des violons plus larges et un chevauchement des distributions, met en évidence les limites du modèle utilisé dans l'article pour distinguer précisément les différences entre les matériaux.

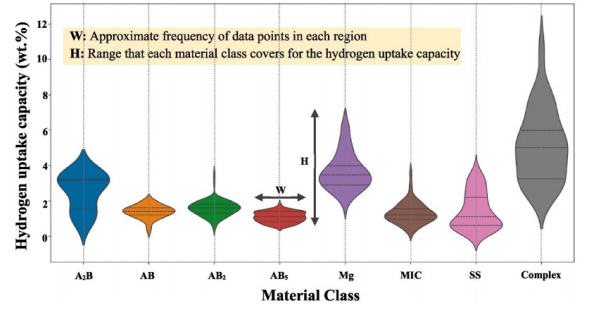


Fig. 4. Violin plot showing a comparison of hydrogen uptake capacity for different material classes.

1217

Figure 13. Distribution de la capacité de stockage d'hydrogène (en %) par classe de matériau obtenue dans l'article étudié.[6]

Cette différence est probablement due à l'utilisation de modèles moins sophistiqués et à l'absence d'enrichissement des données. Bien que cette approche soit adaptée pour des données brutes, elle ne permet pas d'exploiter pleinement les relations complexes entre les propriétés des matériaux (thermodynamiques, atomiques, etc.). Cette écart se justifie par plusieurs raisons. D'abord, l'augmentation des données avec Mendeleev. De plus, le modèle XGBOOST capture les relations non-linéaires entre les variables, ce qui a donné de meilleurs prédictions.

4 Conclusion

L'approche ML s'avère alors plus efficace avec un travail d'enrichissement des données avant le passage à l'entraînement des modèles. Nous avons alors ajouté des propriétés nouvelles ayant des particularités physiques influençant directement le stockage de l'hydrogène, dépassant les approches traditionnelles en termes de cohérence et de fiabilité des prédictions.

Nous proposons alors une methodology à suivre pour optimiser les recherches futures en améliorant le processus de prédiction. Ceci ouvre des portes de développement plus robuste dans la sélectivité du choix des matériaux mis en experimentation. Ainsi, l'énergie durable pourra s'étendre sur divers domaines tout en étant viable sur le côté de recherches technologiques.

5 Perspectives

Cette étude nous amène à de nombreuses perspectives d'amélioration afin de renforcer la précision des résultats obtenus. Tout d'abord, la taille réduite de la base de données comprenant 772 échantillons, limite la capacité des modèles à identifier des relations complexes entre les caractéristiques des matériaux. Un prolongement de cette base, en ajoutant des données supplémentaires ou en fusionnant avec d'autres bases existantes, pourrait améliorer les performances des modèles. Par ailleurs, bien que cette approche ait démontré son efficacité, une validation des résultats par la DFT est essentielle. Ces validations permettront de confirmer les résultats donnés par les modèles.

Un autre défi réside dans l'accès à certaines propriétés clés, telles que les paramètres quantiques ou thermodynamiques, qui restent inaccessibles ou nécessitent des calculs particulièrement coûteux comme ceux de la DFT. Leur intégration, via des techniques avancées d'estimation, enrichirait l'analyse tout en approfondissant la compréhension des matériaux étudiés. De plus, l'étude actuelle s'est principalement focalisée sur la capacité gravimétrique (wt%) des matériaux, mais d'autres propriétés critiques, telles que la stabilité thermique ou la cinétique d'absorption/désorption, mériteraient d'être explorées pour élargir les applications potentielles et offrir des solutions adaptées à des contextes variés.

Enfin, l'optimisation des combinaisons de caractéristiques représente un enjeu majeur. La recherche exhaustive des meilleures combinaisons est actuellement limitée par des contraintes de calcul, nécessitant des temps astronomiques pour explorer toutes les possibilités. L'utilisation de supercalculateurs ou d'infrastructures de calcul avancées pourrait permettre de surmonter cet obstacle, accélérant ainsi l'identification des caractéristiques optimales pour maximiser les performances prédictives. Ces pistes, combinées à l'approfondissement des méthodes utilisées, offrent des opportunités prometteuses pour améliorer les prédictions et favoriser le développement de matériaux optimisés pour le stockage d'hydrogène.

References

- [1] Guoqing Wang, Zongmin Luo, Halefom G. Desta, Mu Chen, Yingchao Dong, and Bin Lin. Ai-driven development of high-performance solid-state hydrogen storage. *Energy Reviews*, 2025. doi:10.1016/j.enrev.2024.100106.
- [2] Manuel Tousignant. L'effet du laminage à froid sur les propriétés de sorption de l'hydrogène du lani5. Master's thesis, Université du Québec à Trois-Rivières, Trois-Rivières, Canada, 2014. Mémoire présenté à l'Université du Québec à Trois-Rivières dans le cadre de la maîtrise en sciences de l'énergie et des matériaux. URL: <https://ntrs.nasa.gov/api/citations/20210018293/downloads/2021%20CEC%20Virtual%20Big%20Tank%20LH2%20DAA%20draft%2007JUL2021.docx.pdf>.
- [3] Mounir Sahli. Synthèse, élaboration et caractérisation des nanocomposites à base de magnésium pour le stockage solide d'hydrogène. 2017.
- [4] Matthew Witman, Mark Allendorf, and Vitalie Stavila. Database for machine learning of hydrogen storage materials properties. <https://zenodo.org/records/10680097>, 2024. Version 0.0.5, consulté le 5 janvier 2025. doi:10.5281/zenodo.10680097.
- [5] Notebook google colab - scientifique, 2025. Consulté le : 5 janvier 2025. URL: <https://colab.research.google.com/drive/14qz00LMNXCCUg2PfJLk8kiaX3m2SJJx?usp=sharing>.
- [6] A. Bhaskar, R. C. Muduli, and P. Kale. Prediction of hydrogen storage in metal hydrides and complex hydrides: A supervised machine learning approach. *International Journal of Hydrogen Energy*, 2025. doi:10.1016/j.ijhydene.2024.12.121.