Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Prediction And Investigation of Transition Metal Oxides for Hydrogen Storage Application via Machine Learning-Assisted Density Functional Theory

Romel Mark Irabon

romelmark.irabon@g.msuiit.edu.ph

Mindanao State University, Iligan Institute of Technology

Rustom Hibone

Mindanao State University, Iligan Institute of Technology

Francis Kirby Burnea

Mindanao State University, Iligan Institute of Technology

Giovanni Paylaga

Mindanao State University, Iligan Institute of Technology

Leo Cristobal Ambolode II

Mindanao State University, Iligan Institute of Technology

Research Article

Keywords: Machine Learning, Domain Adaptation, DFT, Transition Metal Oxides, Hydrogen Storage

Posted Date: March 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9080774/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of Nanoparticle Research on May 18th, 2026. See the published version at https://doi.org/10.1007/s11051-026-06663-5.

Prediction And Investigation of Transition Metal Oxides for Hydrogen Storage Application via Machine Learning-Assisted Density Functional Theory

Romel Mark Irabon[1,3], Rustom Hibone[3], Francis Kirby Burnea[2,3], Giovanni Paylaga[1,3], Leo Cristobal Ambolode II[1,3]

[1]Department of Physics, Mindanao State University – Iligan Institute of Technology, A. Bonifacio Avenue, Tibanga, 9200 Iligan City, Philippines

[2]Department of Chemistry, Mindanao State University – Iligan Institute of Technology, A. Bonifacio Avenue, Tibanga, 9200 Iligan City, Philippines

[3]Center for Nanoscience Research, Premiere Research Institute of Science and Mathematics (PRISM), Mindanao State University – Iligan Institute of Technology, A. Bonifacio Avenue, Tibanga, 9200 Iligan City, Philippines

Email: romelmark.irabon@g.msuiit.edu.ph

Abstract: This study presents a Density Functional Theory (DFT) investigation of transition metal oxides for hydrogen storage, employing machine learning as a supplementary screening tool. Given the scarcity of TMO data, we leverage Unsupervised Domain Adaptation to transfer knowledge from metal hydrides to metal oxides. The Nearest Neighbor Weighting with eXtreme Gradient Boost (NNW-XGB) model demonstrated superior screening performance, ranking promising candidates and predicting hydrogen storage capacities of 3.23 Hwt% for spinel Ca_2TiO_4 and 3.14 Hwt% for perovskite $CaVO_3$. DFT validation using the PBE0 + Grimme's D3 confirmed that both materials maintain structural and electronic stability up to their saturation limits. The comparison shows good agreement for Ca_2TiO_4 (3.03 Hwt% vs. 3.23 Hwt%) but an overestimation for $CaVO_3$ (2.11 Hwt% vs. 3.14 Hwt%), attributed to lattice strain effects not captured by composition-based descriptors. Detailed electronic structure analysis via projected density of states (PDOS) and d-band center calculations revealed that hydrogen adsorption is driven by charge transfer from hydrogen to the B-site transition metal (Ti, V), with the progressive filling of d-states leading to electronic saturation at higher coverages. Compositional trend analysis identified promising combinations of A-site (Mg, Be, Ca, Al) and B-site (Sc, Ti, V, Fe) elements.

Key Words: Machine Learning; Domain Adaptation; DFT; Transition Metal Oxides; Hydrogen Storage

1. Introduction

Hydrogen is highly regarded as a future-proof, cost-efficient, clean fuel. It is the lightest and most abundant element in the universe, with a very high energy density per unit mass (~120 MJ/kg, about three times that of gasoline or diesel), making it a suitable medium for energy storage and harnessing The emerging hydrogen economy has garnered significant attention from political and industrial sectors as a pathway toward environmentally benign energy production and storage technologies [1]. However, the practical realization of hydrogen-based systems faces substantial challenges in storage efficiency, cost, and safety.

The most promising long-term solution is solid-state storage systems. The Hydrogen and Fuel Cell Technologies Office of the U.S. Department of Energy (HFTO-DOE) has established targets for solid-state Hydrogen storage innovations by 2025, including a system gravimetric Hydrogen storage capacity of ~4.5 Hwt%, operational stability for adsorption/desorption between 233 K and 358 K, and reversibility for up to 1500 cycles [2]. While porous materials such as metal-organic frameworks (MOFs), zeolites, and carbon-based structures can achieve capacities up to ~20 Hwt% [3], their reliance on cryogenic temperatures (~77 K) and high pressures limits practical applicability [3]. Metal hydrides, including intermetallic compounds and complex solid solutions, offer improved ambient-condition performance but continue to face challenges in simultaneously achieving fast kinetics and high storage capacity [5].

Emerging as alternatives, Transition Metal Oxides (TMOs) show promise for hydrogen storage. Gebhardt et al. [6] predicted that metal oxide clusters $(MgO)_6$, $(BaO)_6$, and $(BeO)_6$ exhibit hydrogen adsorption energies of 0.15–0.21 eV using PBE0+D3 calculations—well within the reversible storage criteria of 0.16–0.26 eV for moderate temperature operation. Similarly, Goncalves et al. [7] demonstrated, using hybrid DFT calculations, that MnO_2, WO_3, and MoO_2 adsorb hydrogen under storage-relevant conditions, establishing a direct correlation between adsorption energy and material properties such as metal ionic radius, void fraction, and redox potential. Aside from being an effective catalyst for the dissociation/recombination of Hydrogen molecules, oxygen vacancy site defects on TMOs promote greater hydrogen uptake and increased storage capacity [8, 9]. Current literature on oxide-based hydrogen storage has primarily focused on electrochemical applications for Ni/H batteries, with limited

exploration of hydrogen storage under gaseous H_2 pressure and temperature conditions [10, 11, 12, 13].

The fundamental theory underlying DFT calculations begins with the Schrödinger equation for a many-electron system. The Hohenberg-Kohn theorems establish that the ground-state energy is a unique functional of the electron density, while the Kohn-Sham approach reformulates the problem into solving single-particle equations [14, 15]:

$$\left[-\frac{\hbar}{2m}\nabla^2 + V_{ext}(r) + V_H(r) + V_{xc}(r)\right]\psi_i(r) = \varepsilon_i\psi_i(r) \quad (1)$$

where V_{ext} is the external potential from nuclei, V_H is the Hartree potential describing electron-electron Coulomb interactions, and V_{xc} is the exchange-correlation potential that contains all the many-body effects.

The choice of exchange-correlation functional is critical for accurate DFT calculations of metal oxide systems, given the presence of localized d-electrons that govern chemical bonding and reactivity. Standard semi-local functionals such as the Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation (GGA) [16]. While widely used for solid-state systems, they suffer from self-interaction error that leads to over-delocalization of d-electrons, underestimated band gaps, and incorrect descriptions of redox behavior. To address these limitations, hybrid functional, PBE0 [17], incorporate a fraction of exact Hartree-Fock exchange. In PBE0, the exchange-correlation energy combines semi-local GGA with a fraction of nonlocal Hartree-Fock exchange:

$$E_{xc}^{PBE0} = E_{xc}^{PBE} + 0.25(E_{HF} - E_{xc}^{PBE}) \quad (2)$$

Despite the accuracy of DFT methods, the vast compositional space of TMOs, encompassing perovskites (ABO_3), spinels (AB_2O_4), pyrochlores, and more complex structures, makes exhaustive first-principles screening computationally prohibitive. To address this challenge, machine learning approaches have emerged as efficient preliminary screening tools that can guide subsequent DFT investigations toward the most promising candidates [18, 19, 20]. However, the scarcity of experimental and computational hydrogen storage data for TMOs poses a significant challenge for training predictive models, particularly regarding their ability to generalize to new material classes.

Domain adaptation (DA), a subset of transfer learning, addresses this challenge by enabling knowledge transfer from a source material dataset abundantly found in literature, to a material class with unknown properties by minimizing distribution divergence [21]. For scenarios involving unlabeled materials, Unsupervised Domain Adaptation (UDA) proves particularly valuable [22, 23]. UDA can also accommodate large target datasets relative to the source, relaxing transferability restrictions and enabling discovery of unexplored materials [22, 23].

In this study, we adopt an ML-assisted DFT approach to predict the hydrogen storage properties of transition metal oxides using compositional descriptors. Recognizing the abundance of metal hydride data and the scarcity of TMO data, we train machine learning models on metal hydrides and leverage UDA to facilitate knowledge transfer to TMO systems. Promising TMO candidates are ranked based on ML-predicted hydrogen storage capacity and subsequently validated through explicit DFT calculations using the PBE0 hybrid functional with Grimme's D3 dispersion correction. Finally, the hydrogen storage mechanisms, energetics, and stability of TMOs at the electronic structure level were also presented.

2. Methodology

2.1. Data Acquisition and Featurization

The metal hydride dataset was obtained and expanded from the publicly available database provided by Witman et al. [24, 25], which contains 462 multiclass (AB, A_2B, AB_5, Mg-Based, Intermetallic Compounds, and Solid Solutions). The dataset, including an additional 63 metal hydrides compiled from recent literature, was preprocessed to remove missing and duplicate entries, yielding a final total of 432 unique metal hydrides. On the other hand, the Transition Metal Oxides dataset has a total of 1455 metal oxides consist of 60 TMOs queried from a web-based open-source computational materials database called Materials Project [26], 369 ABO_3 perovskite oxides that are predicted stable by Emery and Wolverton (2017) [27], 451 stable AB_2O_4 spinel oxides by the groups of Li et al. (2020) [28], and 575 complex metal oxides consisting of double perovskite oxides and pyrochlores from the dataset of Baldassari et al. [29]. The datasets were featurized using the Material-Agnostic Platform for Informatics and Exploration (MAGPIE) to generate composition-based descriptors [30]. The algorithm generates an array of 132 composition-based feature vectors that incorporate stoichiometric attributes, elemental properties, electronic-structure descriptors, and ionic-compound

characteristics derived solely from the chemical formula. These feature vectors were constructed using the Composition-Based Feature Vector (CBFV) Python module [31] . From 132 MAGPIE default features, it was reduced to 24, containing the min, max, std, and avg of the following, which we believed were Hydrogen storage-relevant attributes: covalent radius, electronegativity, number of filled valence electrons, atomic weight, Space Group Number, and Ground-State Volume per atom. The correlations across the feature space in the metal hydrides dataset were capped at 0.95.

2.2.Base Model Training and Prediction

Machine learning regression models from the SKLearn Python module, Extreme Gradient Boost (Boost), Decision Tree, Random Forest, ADABoost, and Linear Support Vector Regressor (Linear SVR) were considered as base estimators and compared. The base estimator was initially trained on the source domain, which is the metal hydride dataset. The gravimetric Hydrogen storage capacity Hwt% and 24-dimensional features were labelled as y, and X, respectively, and were split into an 80% training dataset (*X_train, y_train*) and 20% test dataset (*X_test, y_test*). Finally, the training dataset was fitted to the machine learning models, and hyperparameter tuning was performed via 10-fold cross-validation (CV=10) using *GridSearchCV.* Metric scores such as R^2, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) were used to quantify the overall accuracy of the base estimators/models to determine the optimal base estimator.

2.3. Unsupervised Domain Adaptation

The MAGPIE-generated features were still used and assigned as the target domain input variables (Xt). Similarly, the feature input space and the Hydrogen storage capacity (Hwt%) of the metal hydrides dataset were designated as the source domain input variables (Xs) and corresponding output labels (ys), respectively. The best-performing base estimator, along with the generated working dataset for metal hydrides and metal oxides, was used as input to unsupervised domain adaptation algorithms.

We conducted preliminary tests limited to domain adaptation models in the Python-based Awesome Domain Adaptation Python (ADAPT) package module [32] that supports unsupervised learning. As parameter-based approaches primarily require supervised learning, we selected feature-based and instance-based models for the pre-evaluation. Initial findings

based on source-domain predictions revealed that instance-based models outperformed feature-based models, apparently due to their robustness in handling unlabeled data in a materials science setting [22]. Three instance-based UDAs were tested further in this study: the Kullback-Leibler Importance Estimation Procedure (KLIEP) [33], Unconstrained Least-Squares Importance Fitting (ULSIF) [34], and Nearest Neighbor Weighting (NNW) [35].

KLIEP aims to minimize the Kullback-Leibler divergence between the source and target distributions. ULSIF minimizes the squared error and loss, whereas NNW uses a 1-Nearest Neighbor classifier to determine sample weights. Since KLIEP and ULSIF have a built-in cross-validation algorithm, a loop-based automated search was conducted for hyperparameter tuning. To evaluate the transferability performance of the UDA approach, we implemented both qualitative and quantitative assessments. For visual analysis, we compared the source and target distributions using kernel density estimation (KDE) on principal component analysis (PCA)-transformed feature spaces. Quantitatively, we adopted two complementary metrics: (1) Source Accuracy, and (2) Maximum Mean Discrepancy (MMD). Source Accuracy [22] is measured by transferring the sample weights predicted by the UDA to the base estimator and then predicting on the same test set to compare metric results, ensuring that no information is lost when transferring knowledge from the source to the target domain. Maximum Mean Discrepancy or MMD [36] It is a statistical measure used to assess the similarity and differences between two domain feature spaces based on the source and target distribution's mean and covariance matrices. The optimal UDA model, based on analysis, was utilized to predict the Hydrogen storage capacity of the transition metal oxides (TMOs).

2.4. Transition Metal Oxide (TMO) Selection

As a proof-of-concept of the model's predictive power, we rank TMOs by their Hydrogen Storage Capacity. We chose two materials to represent spinel and perovskite oxides, namely Ca_2TiO_4 and $CaVO_3$, to validate their Hydrogen adsorption properties using DFT. The crystal structures were obtained from the Open Quantum Materials Database (OQMD) [37]. The selection of the polymorph for each material was not strictly constrained, as this study focuses primarily on composition-based predictions rather than detailed polymorph-specific effects. The spinel oxide Ca_2TiO_4 (structure ID 5222550), from the family of Ruddlesden–Popper (RP) series, exhibits a rich polymorphic landscape due to octahedral tilting distortions [38]. Among its polymorphs, the orthorhombic phase (Pbam), a more stable, distorted-symmetry phase derived from its common I4/mmm structure, is considered in this study. On the other hand,

perovskite oxide $CaVO_3$ (structure ID 2005724) has been experimentally observed to exist in orthorhombic phase (Pnma) as its most stable form [39] and is also considered in this study.

1.5. Computational Details.

Quantum Mechanical simulation was utilized in all calculations based on Quantum Espresso's Density Functional Theory [40] to examine the electronic properties and structural stability of the two TMOs. The Exchange-Correlation Approximation Functional Perdew-Burke-Ernzerhof (PBE0) [17] was used to model the electronic interactions within the system with the empirical Grimme's D3 [41] correction to facilitate Van der Waals interaction between the atomic Hydrogen and the surface of the material.

Structure optimization and Self-Consistent Field (SCF) calculations were conducted using the Garrity-Bennett-Rabe-Vanderbilt (GBRV) ultrasoft pseudopotential [42] for its accuracy and efficiency, balanced with a 40 Ryd (~544 eV) plane-wave energy cutoff, as recommended by the proponents. The convergence criterion is set until the energy and force have been minimized to 10^4 Ryd ($\sim 1.36 \times 10^3\ eV$) and 10^3 Ryd/Bohr ($\sim 0.026\ eV/$ Å), respectively. The Brillouin zone of the relaxed atomic positions for both pristine and hydrogen-adsorbed cases was sampled using a Gamma-point-only k-point mesh, employed with Gaussian broadening of 0.02 Ryd (~0.27 eV) to ensure convergence for all calculations.

For the Hydrogen adsorption energy calculations, a 2 × 1 × 2 supercell containing 56 atoms for the pristine Ca_2TiO_4 and a 2 × 1 × 1 supercell with 40 atoms for $CaVO_3$ were constructed with a 10 Å vacuum applied along the Z-direction. The Ca–O rocksalt-terminated surface along the plane [001], where the stacking is terminated, was chosen as the Hydrogen adsorption site for both materials due to its reported stability [43] while the lattice atomic configuration was calculated, and the most energetically favorable one was identified.

Hydrogen atoms were initially loaded and saturated on top of the oxygen uncoordinated sites with an O-H distance of 1 Å before loading other sites (**Fig. 1**). The total energy of the system for the pristine and surface Hydrogen adsorbed cases, namely Ca2TiO4Hx (x = 0, 4, 6, 8) and CaVO3Hx (x = 0, 3, 6, 9), was determined using a denser 2 × 2 × 2 Monkhorst-Pack k-point grid. The calculation of the Hydrogen adsorption energy ΔE_{ads}^{H} is given by the formula [18],

$$\Delta E_{ads}^{H} = \frac{E_{TMO+H} - \left(E_{TMO} + \frac{1}{2}E_{H2}\right)}{nH} \quad (3)$$

213

214 Where E_{TMO} and E_{TMO+H} These are the DFT calculated energies of the pristine case and surface
215 H-adsorbed case of the material, respectively. A separate reference energy calculation of E_{H2}
216 for the H_2 molecule was carried out at the same level of DFT theory under 12 Å × 12 Å × 12

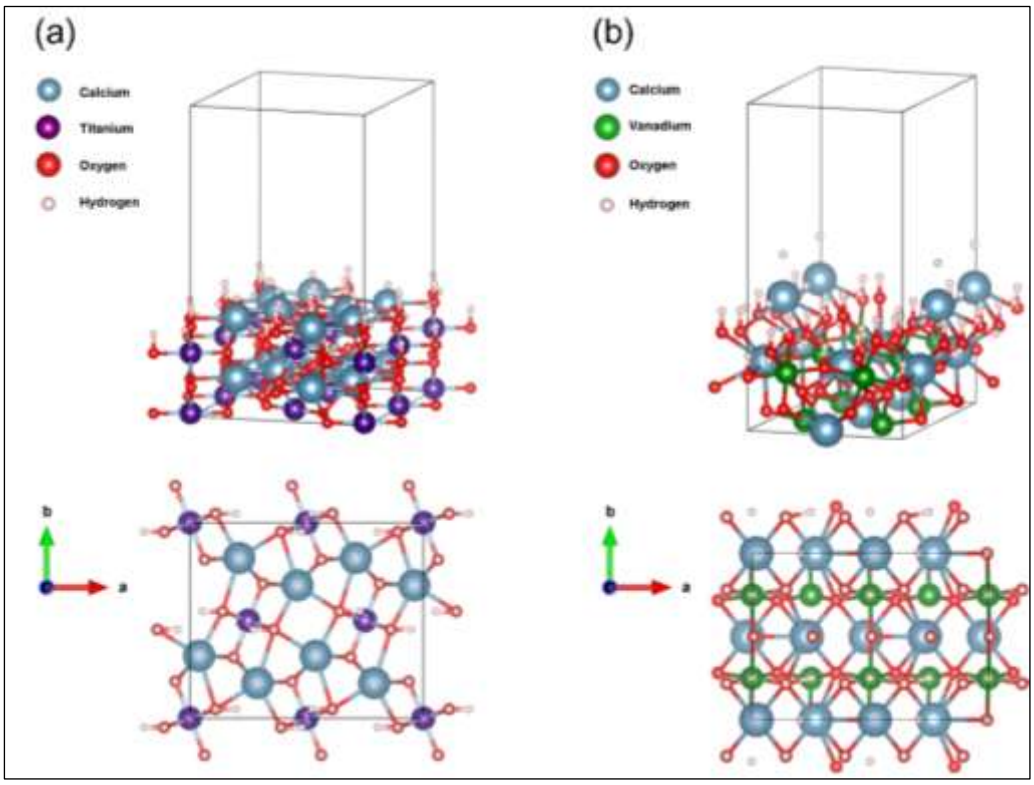


Fig. 1 Initial structure configuration of Hydrogen loaded (a) $Ca_2TiO_4H_4$, and (b) $CaVO_3H_3$.

217 Å vacuum conditions, using a dense 6 × 6 × 6 k-point grid. On the other hand, the gravimetric
218 Hydrogen storage capacity $Hwt\%$ can be resolved using the equation [11],

$$Hwt\% = \frac{n\, m_H}{m_{mat} + m_H} \quad (4)$$

219

220 where m_H is the mass of the Hydrogen and m_{mat} is the mass of the material. Finally, Bader
221 charge analysis with the code provided by the Henkelman group. [44] was implemented using
222 the Projector Augmented Wave (PAW) method to accurately represent the full electronic charge
223 density and the bonding nature of the atoms.

Lastly, we performed projected density of states (PDOS) analysis and calculated the d-band center for the B-site transition metal (Ti, V). The PDOS was obtained from non-self-consistent-field (NSCF) calculations, with H2-adsorbed and pristine cases as references. The d-band center (ε_d) was computed as the first moment of the projected d-states relative to the Fermi level (E_f) using this equation:

$$\epsilon_d = \frac{\int_{-\infty}^{\infty}(E - E_f)\ \rho_d\ (E)dE}{\int_{-\infty}^{\infty}\rho_d(E)dE} \quad (5)$$

where $\rho_d(E)$ is the d-orbital projected density of states.

3. Results and Discussion

Table 1. Performance metrics of the five machine learning models for the metal hydride dataset.

	R^2	MAE	RMSE
Linear Support Vector	0.801	0.337	0.464
ADABoost	0.907	0.214	0.317
Decision Tree	0.858	0.267	0.393
Random Forest	0.924	0.207	0.288
XGBoost	**0.936**	**0.190**	**0.263**

3.1. Machine Learning Model Performance and Domain Adaptation

Five regression models—Linear Support Vector Regression, Decision Tree, Random Forest, AdaBoost, and Extreme Gradient Boosting (XGBoost)—were trained on the metal hydrides dataset (**Fig. S1**). To eliminate sampling bias and overfitting, we again perform 10 cross-validations to tune the parameters via *GridSearchCV*. Performance metrics for training and testing are reported in Table 1. XGBoost outperformed other models, achieving $R^2 = 0.936$, MAE = 0.190, and RMSE = 0.263, as shown in **Fig. 2a**. Optimal hyperparameters identified through hyperparameter tuning include the following: learning rate = 0.7, n_estimators = 1250, max_depth = 3, colsample_bytree = 0.4, and colsample_bynode = 0.3. This demonstrates the robustness of XGBoost in handling dataset diversity and limited sample sizes. As a tree-based gradient boosting model, XGBoost effectively handles complex regression tasks for small

datasets while capturing intricate patterns and feature importance. Its L1 (Lasso) and L2 (Ridge) regularization mitigate overfitting, which is essential for complex decision boundaries in high-dimensional feature spaces [45]. For these reasons, XGBoost was selected as the base estimator for unsupervised domain adaptation (UDA).

We also conducted a feature importance analysis using XGBoost's built-in Python module, which uses a gain-based importance method**.** Feature importance, in **Fig. 2b**, revealed the top predictors for Hydrogen storage capacity: Minimum Number of Valence Electrons, Minimum Atomic Weight, Average Number of Valence Electrons, Average Electronegativity, and Minimum Space Group Number. This aligns with literature where Valence Electron Concentration (VEC) correlates with Hydrogen storage performance [46]. Furthermore, doping with low-valence-electron elements enhances capacity, while composition-based metrics (e.g., electronegativity distributions) critically influence Hydrogen adsorption. [47]. These reported consistencies with the literature substantiate the model's interpretability and reinforce confidence in its predictions.

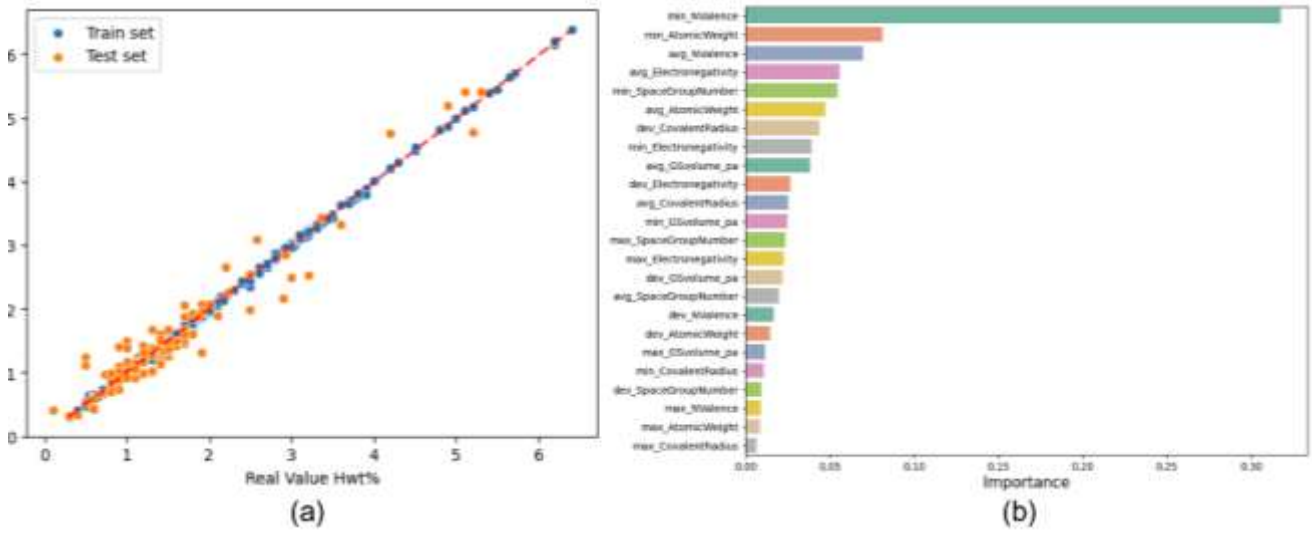


Fig. 2 (a) Predicted vs Real Value of maximum Hwt% in metal hydrides dataset using XGBoost Regression, and (b) gain-based feature importance result from model training.

We evaluated both instance-based and feature-based unsupervised domain adaptation (UDA) algorithms using XGBoost as the base estimator. The instance-based methods consistently outperformed the feature-based approaches (**Fig. S2-S3**). Therefore, we selected three instance-based algorithms, namely KLIEP XGB, ULSIF XGB, and NNW XGB, for further testing. For a machine learning model trained with the DA algorithm, its prediction

performance on the source domain must not degrade, even when trained on a different chemical space. We confirmed the distribution discrepancy between the source (metal hydrides) and target (metal oxides) domains using Principal Component Analysis with Kernel Density Estimation (PCA-KDE), a method adopted from the literature [48, 49, 50]. A visual comparison of the unweighted and weighted PCA-KDE plot for the three UDA models is shown in **Fig. 3**. The first two principal components (PC1–PC2) accounted for 70.6% and 75.9% of the cumulative variance in metal hydrides and metal oxides, respectively.

Effective domain adaptation requires minimizing the discrepancy between domains. The alignment is achieved by projecting the source domain subspace onto the target subspace until their distributions align [46,47]. Among the three UDA algorithms evaluated, NNW+XGB demonstrated the most effective domain alignment. In contrast, both KLIEP+XGB and ULSIF+XGB exhibited susceptibility to outliers, which may introduce errors in predicting Hydrogen storage properties for TMO systems. While that is the case, NNW's weighting mitigated this in our dataset.

Quantitative evaluation revealed a significant initial distributional discrepancy between the source (metal hydrides) and target (metal oxides) domains, with a Maximum Mean Discrepancy (MMD) of 1.384. Note that for ideal domain datasets, an MMD approaching zero indicates perfect distributional alignment. All UDA methods demonstrated improved domain adaptation. Among these, KLIEP-XGB exhibited the least improvement, achieving a source prediction R^2 of 0.937 and an MMD value of 0.976. In contrast, ULSIF-XGB and NNW-XGB showed more substantial improvements in source-domain predictions, with source-domain predictions of 0.941 and 0.940, respectively. NNW-XGB outperformed all models in reducing distribution shift, achieving a post-adaptation MMD of 0.751 compared to 1.123 for ULSIF-XGB.

Hu et al. (2023) [22] observed the high performance of unsupervised domain adaptation (UDA) methods—such as KLIEP, uLSIF, and NNW—in machine learning models for material property prediction. In their paper, they highlighted that instance-based domain adaptation (DA) approaches, particularly NNW, outperformed other models in material property prediction across large chemical spaces. Instance-based methods adjust weights to individual data points, which is particularly effective in materials science, where datasets often exhibit unique regularities.

Having identified NNW+XGB as the most effective UDA approach based on reduced distribution discrepancies and improved metrics, we applied the model to predict Hydrogen storage capacities of TMOs. To validate these predictions, we first ranked the TMOs by predicted storage capacity, then selected two representative TMOs (Ca_2TiO_4, spinel; $CaVO_3$, perovskite) for detailed computational analysis. The top 10 transition metal oxides were shown in

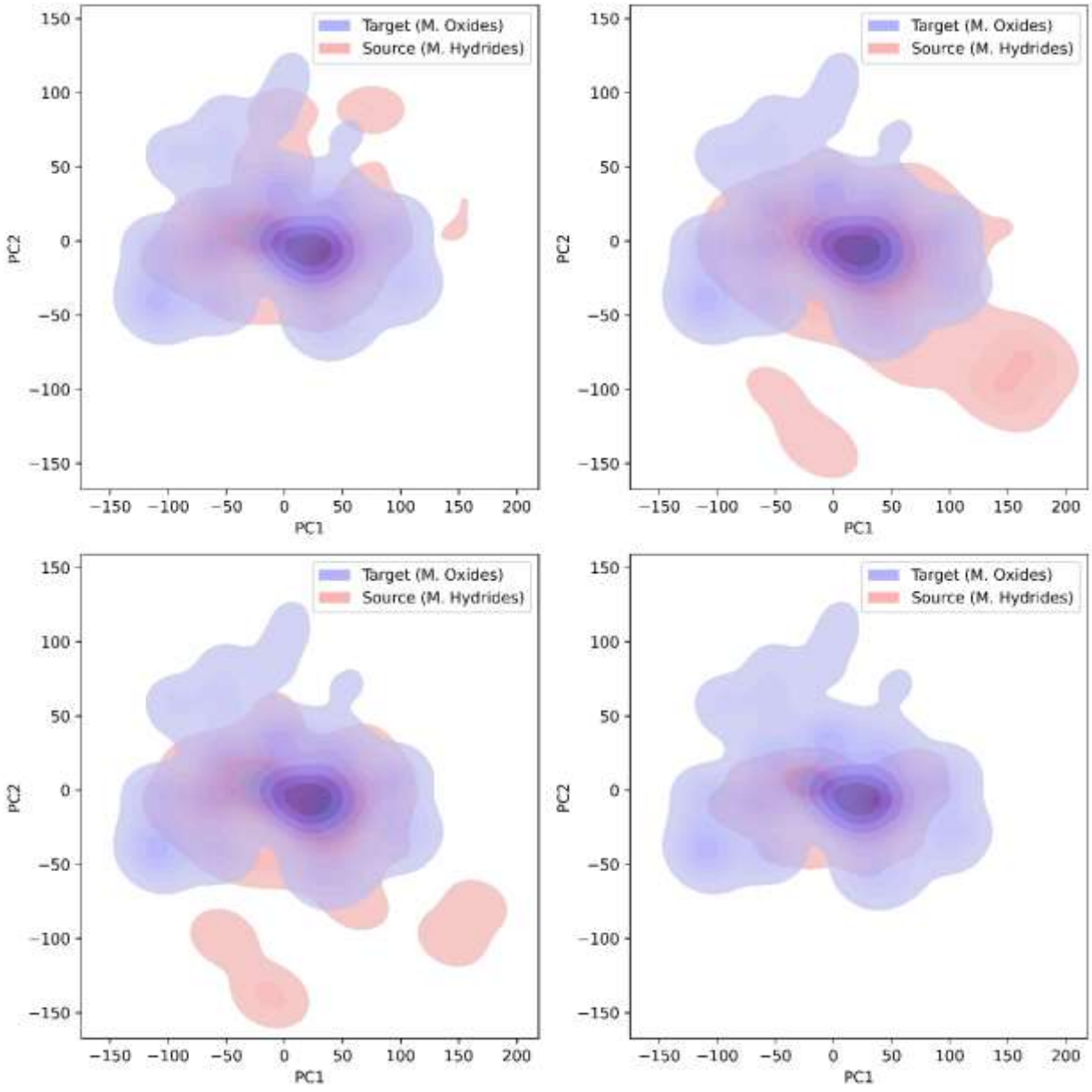


Fig. 3 KDE-PCA plot of the unweighted and UDA-weighted source data.

Table *2*.

Table 2. The table shows the top 10 transition metal oxides based on gravimetric storage capacity in Hwt%. Selection criteria: one (1) perovskite oxide ABO_3, and one (1) spinel oxide A_2BO_4, where B must be a transition metal. $CaTiO_3$ is an option; however, there is existing literature. Therefore, we selected $CaVO_3$ and Ca_2TiO_4 as the materials to confirm H Storage capacity via the DFT PBE+D3 approach.

Transition Metal Oxides	NNW+XGB Predicted Hwt%
Ca_2TiMnO_6	3.301070
$CaSc_2O_4$	3.280571
$TiCa_2O_4$	3.225121
$CaTiO_3$	3.213571
$CaVO_3$	3.138927
$MgSc_2O_4$	2.877885
$CaFeO_3$	2.762358
Ca_2VFeO_6	2.744353
Sc_2O_3	2.676759
$ScSc_2O_4$	2.447782

3.2. Hydrogen Adsorption Energetics and Charge Transfer Analysis

We inspected the structures and corresponding energies of the relaxed configurations of $Ca_2TiO_4H_x$ (x = 0, 4, 6, 8) and $CaVO_3H_x$ (x = 0, 3, 6, 9), where x = 0 corresponds to the pristine structures and higher x values denote the number of Hydrogen atoms adsorbed on the Transition Metal Oxide surfaces. All hydrogenated systems are energetically stable, with negative Hydrogen adsorption energies indicating exothermic processes. For hydrogenated Ca_2TiO_4, the calculated adsorption energies per H atom were -0.203 eV, -0.190 eV, and -0.155 eV for H_4, H_6, and H_8, respectively. For hydrogenated $CaVO_3$, the values were -0.754 eV (H_3), -0.464 eV (H_6), and -0.320 eV (H_9). The progressive weakening of E_{ads} and structural distortion (**Fig. S4-S5**) with increasing coverage suggests the onset of electronic saturation and H-H repulsive interactions.

To understand its origin, we performed Bader charge analysis. The violin plots of the atomic charge redistribution ΔQ with increasing surface Hydrogen coverages for $Ca_2TiO_4H_x$ (x = 0, 4, 6, 8) and $CaVO_3H_x$ (x = 0, 3, 6, 9) are illustrated in **Fig. 4**. A violin plot is a powerful tool for visualizing data distributions, combining the information of a box plot with a kernel density plot. Each violin represents the full probability density of ΔQ values for a specific element (Ti, Ca, O, H) at a given hydrogenation state of the TMOs. The horizontal width of the violin represents the concentration of atoms in a specific ΔQ, while the vertical spread represents the range and distribution of charge differences.

In Ca_2TiO_4 **Fig. 4a**, Titanium acts as the main redox-active site, with ΔQ values averaging +0.285 e^- (H_4), +0.198 e^- (H_6), and +0.416 e^- (H_8), consistent with $Ti^{4+} \rightarrow Ti^{3+}$ reduction. On the other hand, Calcium shows small electron losses (−0.028 to −0.033 e^-), while Oxygen remains nearly unchanged, averaging ΔQ within ±0.04 e^-. Hydrogen exhibits a negative overall ΔQ, characteristic of protonic donation, although a broadened, bimodal distribution at H_8 may suggest a mixture of protonic and hydridic states. At high coverage, broader distributions and bimodal ΔQ for Ti and O were also observed, suggesting charge disproportionation and local electronic frustration. These results align with weakening adsorption energies as surface Hydrogen coverage increases, consistent with electronic saturation and H–H repulsion.

A similar progression and a more pronounced progression were observed in $CaVO_3$ (**Fig. 4b**), where Vanadium shows the most pronounced charge redistribution. V gains electrons steadily across coverages, from +0.362 e^- in H_3 to +0.667 e^- in H_9, marking a reduction from V^{5+} toward V^{4+}/V^{3+}. O gains electrons (+0.273 e^- at H_3 to +0.358 e^- at H_9), indicative of O–H bond formation, while Ca remains nearly invariant (~+0.02 e^-). Hydrogen acts as a donor overall, but distributions broaden substantially at H_6 and H_9, with a bimodal spread for both V and O signaling charge heterogeneity.

These Bader analyses substantiate UDA predictions, demonstrating that both TMOs support low-to-moderate H capacities driven by B-site reduction, yet are limited by electronic saturation and structural strain at high coverages. DFT simulations indicate that the optimal

hydrogen storage capacities are Ca2TiO4H6 (3.03 Hwt%) and CaVO3H3 (2.11 Hwt%), compared with machine-learning predictions of 3.23 Hwt% and 3.14 Hwt%, respectively.

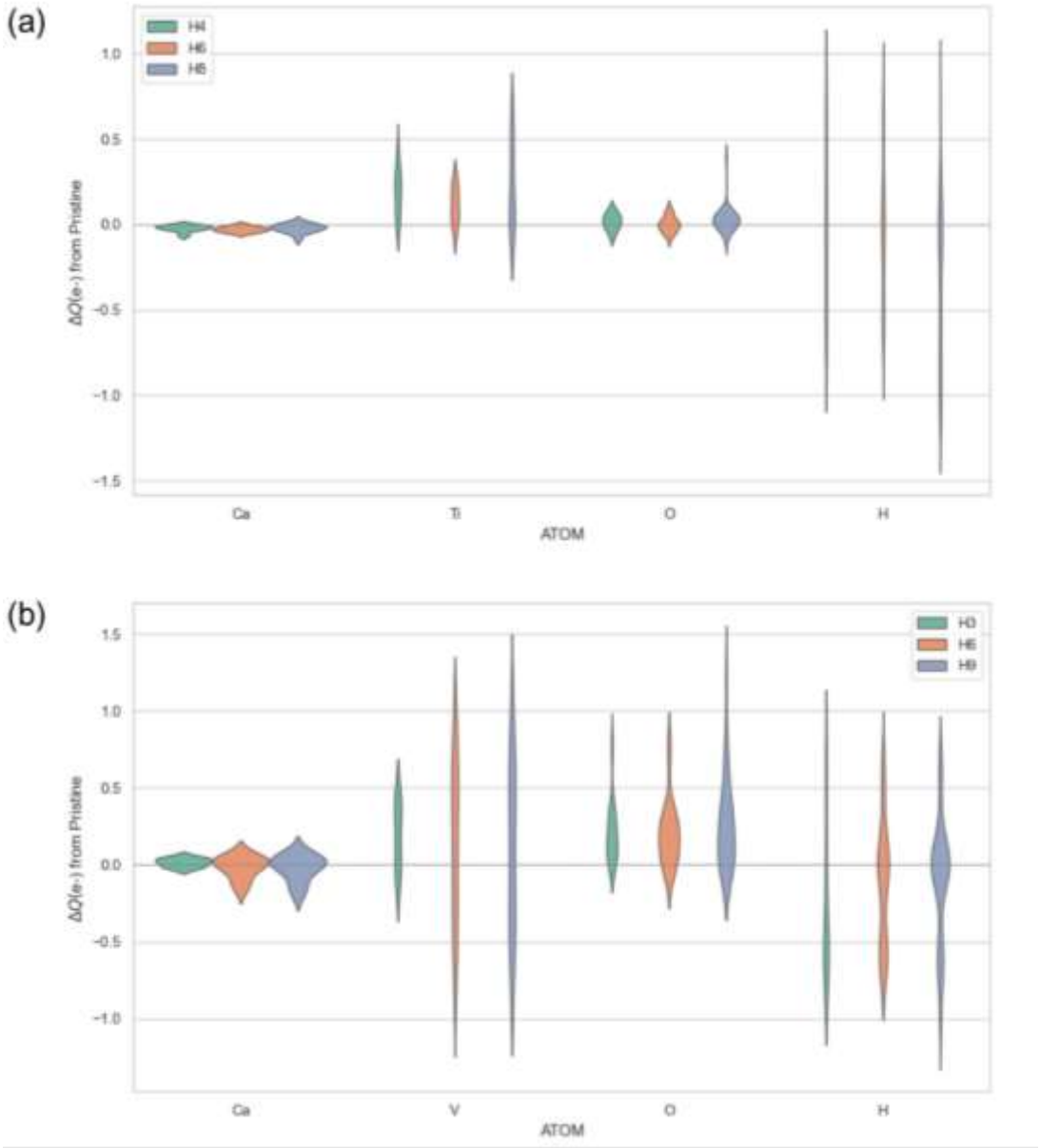


Fig. 4 Violin plot of the atomic charge transfer ΔQ in (a) $Ca_2TiO_4H_x$ (x=0,4,6,8) and (b) $CaVO_3H_x$ (x=0,3,6,9).

3.3. Electronic Structure Analysis: Projected Density of States and d-Band Center

To further elucidate the electronic origin of the hydrogen adsorption behavior and the observed saturation, we conducted a detailed analysis of the projected density of states (PDOS). We calculated the d-band center shift relative to the Fermi Energy for the B-site transition metals (**Fig. S6-S7**). The d-band center is a key descriptor of a transition metal's reactivity; a

downward shift relative to the Fermi level typically indicates stronger interaction with adsorbates due to greater filling of antibonding states, thereby stabilizing the adsorption complex [53].

The pristine Ca_2TiO_4 [001] (**Fig. 5-left**) exhibits a valence band dominated by strong O 2p - Ti 3d hybridization from approximately -8 eV to the Fermi level. Ca states show minimal contribution, confirming its role as a structural buffer, consistent with the Bader analysis. The Ti d-band center is located at +0.022 eV relative to E_f. This position, essentially at the Fermi level, indicates a high density of unoccupied or partially occupied d-states available for accepting electron density from an adsorbate. Upon hydrogen molecule adsorption (**Fig. 5-right**), a distinct H 1s peak emerges at approximately -6 eV, which exhibits strong overlap with the O 2p band. This signifies the formation of a stable O-H bond, where the H 1s electron is donated into the O 2p-dominated bonding states. Crucially, the Ti d-band center shifts substantially downward to -0.892 eV. This significant shift of -0.914 eV is a direct consequence of electron transfer from the adsorbed hydrogen to the Ti sites. The donated electrons occupy previously empty Ti d states, effectively reducing Ti^{4+} to Ti^{3+}. The position of the new d-band center, below E_f explains the saturation behavior at higher coverages; once these lower-energy states are filled, additional electrons must occupy higher-energy, antibonding states, which are energetically unfavorable and lead to the observed charge heterogeneity in the Bader analysis.

On the other hand, the pristine $CaVO_3$ [001] surface is metallic in nature, as shown in **Fig. 6-left**. The V d-band center is located at -1.053 eV, already below E_f, indicating a partially filled d-band and a high intrinsic reactivity. Hydrogen adsorption in $CaVO_3$ (**Fig. 6-right)**, also induces electronic reorganization. New hybridized states appear throughout the valence band, with distinct H 1s contributions overlapping with both O 2p and V 3d states, confirming the formation of both O-H and V-H bonding interactions. The V d-band center undergoes a dramatic further downward shift to -1.502 eV. This additional -0.449 eV shift is significantly larger than that observed in the spinel oxide, reflecting the stronger electron-accepting capability of vanadium. The final d-band center is positioned deep below E_f, indicates a high degree of filling of the V d-states. This electronic saturation explains why further hydrogen uptake is severely limited. The metal center is already highly reduced, and additional charge leads to the pronounced electronic frustration and charge heterogeneity observed in the Bader analysis.

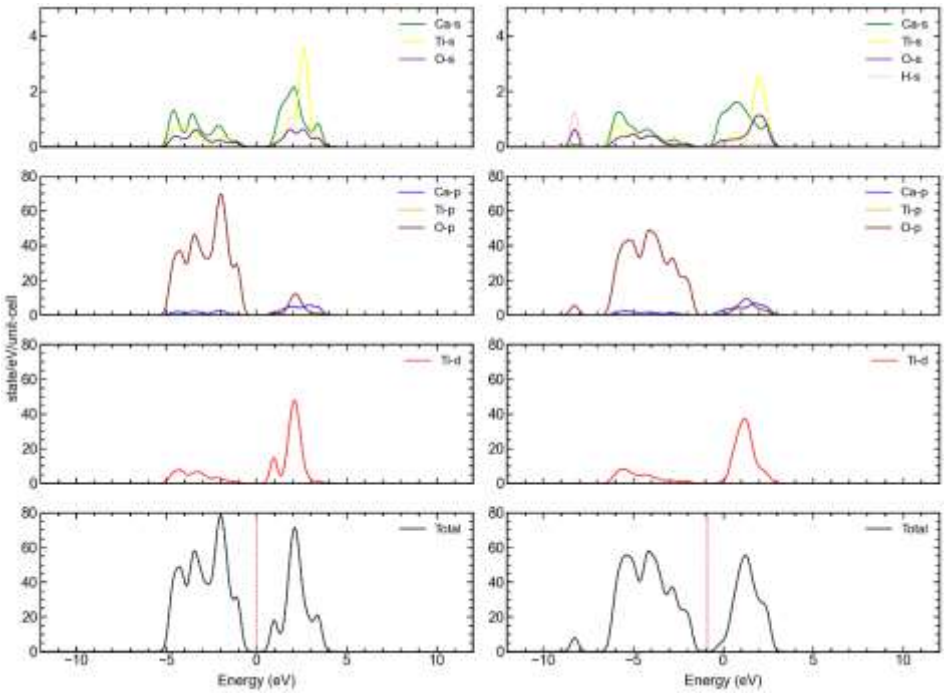


Fig. 5 Projected DOS of pristine (left) and H-adsorbed Ca_2TiO_4 (right). The vertical dashed line indicates the d-band center (ε_d) of Ti.

386

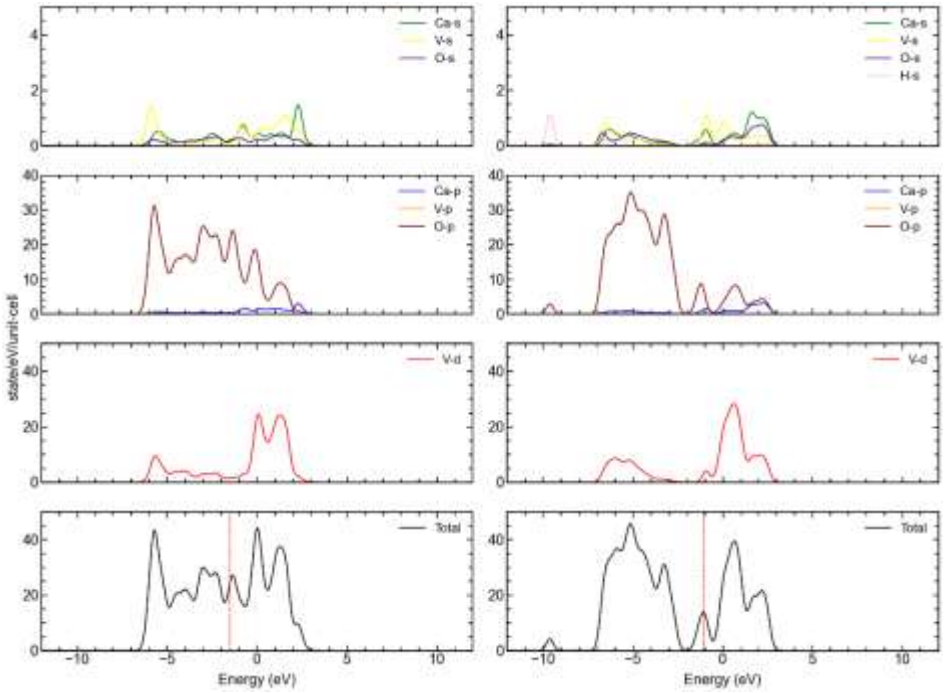


Fig. 6 Projected DOS of pristine (left) and H-adsorbed Ca2TiO4 (right). The vertical dashed line indicates the d-band center (ε_d) of V.

3.4. Comparison of Predicted and Calculated Storage Capacities

DFT simulation results have shown that the optimal case for Hydrogen storage capacity is $Ca_2TiO_4H_6$ (3.03 Hwt%) and $CaVO_3H_3$ (2.11 Hwt%), compared with machine learning predictions of 3.23 Hwt% and 3.14 Hwt%, respectively. In **This discrepancy** is likely due to the limitations imposed by composition-based descriptors, which the model's current framework does not fully account for. Furthermore, intermediate hydrogenation states with capacities between the H_3 and H_6 cases were not investigated here and may still represent stable configurations, which warrants future exploration.

Table 3, we used the model to predict the Hwt% of certain TMOs reported in the literature and compared the results with their reported values. The close agreement for Ca_2TiO_4 underscores the robustness of the NNW-XGB model in capturing Hydrogen uptake trends; however, the underestimation of CaVO3 suggests that the model cannot capture trends in systems where structural instability and lattice strain effects emerge. This discrepancy is likely due to the limitations imposed by composition-based descriptors, which the model's current framework does not fully account for. Furthermore, intermediate hydrogenation states with capacities between the H_3 and H_6 cases were not investigated here and may still represent stable configurations, which warrants future exploration.

Table 3. NNW-XGB model and DFT Predicted Hydrogen storage capacity (Hwt%) of certain TMOs.

TMO	ML Predicted (Hwt%)	DFT Predicted (Hwt%)	Source
Ca_2TiO_4	3.23	3.03	This paper
$CaVO_3$	3.14	2.11	This paper
$BaYO_3$	1.00	1.09	[54]
$CaTiO_3$	3.21	4.27	[52]

In comparison with the U.S. DOE target of 4.5 Hwt%, the predicted storage capacities of pristine TMOs fall short of the standard. However, these values could be represented as lower-bound estimates when compared to real-world TMO systems [55, 56]. Based on the rankings of predicted storage capacities, we identified potential elements to guide future synthesis and optimization strategies. Promising B-site transition metals include Sc, Ti, V, and Fe, while viable A-site elements are Mg, Be, Ca, and Al. These compositional trends could

offer a clear direction for designing defect-tuned or doped TMOs, where vacancy engineering may enable Hydrogen storage capacities beyond the projected standard.

4. Conclusion

In this work, we have successfully demonstrated a machine-learning-assisted DFT framework to accelerate the discovery of metal oxides for solid-state Hydrogen storage. Facing the challenge of scarce data for metal oxides, we employed Unsupervised Domain Adaptation (UDA) to effectively transfer knowledge from the well-established chemical space of metal hydrides. Among the tested algorithms, the Nearest Neighbor Weighting (NNW) method coupled with XGBoost emerged as the most robust model, significantly reducing the distributional discrepancy between the source and target domains and yielding reliable predictions.

Feature importance analysis revealed that low valence electron concentration is a critical descriptor for Hydrogen storage capacity, aligning with established physical principles and lending interpretability to our model.

We reported the top 10 promising transition metal oxides and then investigated the spinel Ca_2TiO_4 and perovskite $CaVO_3$ in detail for DFT validation. Our DFT calculations confirmed the feasibility of Hydrogen adsorption on these materials but also revealed their practical limitations. Bader charge analysis provided deep insight into the adsorption mechanism, identifying the reduction of the B-site transition metal (Ti, V) as the primary driver for Hydrogen uptake. This was further corroborated by detailed projected density of states (PDOS) and d-band center analysis, which showed that hydrogen adsorption induces a significant downward shift in the d-band center of Ti and V. This analysis revealed that electronic saturation, via progressive filling of available d-states, is the fundamental cause of the capacity limits observed at higher coverages, particularly in the perovskite oxide $CaVO_3$.

The machine learning predictions (3.23 Hwt% for Ca_2TiO_4, 3.14 Hwt% for $CaVO_3$) showed a close agreement with DFT for the spinel $Ca_2TiO_4H_6$ (3.03 Hwt%). Still, they overestimated the capacity for the perovskite $CaVO_3H_3$ (2.11 Hwt%), highlighting the model's current limitation in capturing material-specific instability arising from electronic saturation and lattice strain.

Addressing these challenges requires future work on doped variants and heterostructures to mitigate saturation limits (e.g., by modifying the d-band center), and on

developing stability-aware descriptors to improve the predictive fidelity of data-driven approaches.

Acknowledgement

The authors are grateful to acknowledge the support provided by the Office of the Vice Chancellor for Research and Enterprise (OVCRE) of Mindanao State University-Iligan Institute of Technology (MSU-IIT) (S.O. No. 00231-IIT S. of 2024) and the Accelerated Science and Technology Human Resource Development Program (ASTHRDP) of the Department of Science and Technology (DOST), Republic of the Philippines. We also extend our sincere thanks to Prof. Johniel E. Babiera from the Mathematics & Statistics Department of the College of Science and Mathematics, for sharing his insights in this study. Computational resources were provided by the Computing and Archiving Research Environment (COARE) of the Department of Science and Technology–Advanced Science and Technology Institute (DOST-ASTI), which the authors gratefully recognized.

Author Contributions

Romel Mark Irabon – Conceptualization, formal analysis, writing final draft; Rustom R. Hibone – investigation, writing, formal analysis; Francis Kirby B. Burnea – Principal Investigator, supervision, review; Giovanni J. Paylaga – supervision, review; Leo Cristobal C. Ambolode II – supervision, conceptualization, review

Data Availability

Supporting Information for this paper is available from the journal's website. Additionally, the trained machine learning model, source code, and dataset is publicly accessible via this link: https://zenodo.org/records/18932805?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6Ijc5YmY5YWZmLWNlOTctNDg2Yi05YmJkLWIyNzBjNjY5ZTU5ZiIsImRhdGEiOnt9LCJyYW5kb20iOiIxNmQ0YjkzMDQ4MzNiN2E5NzkyYTg4YTE2YzE3MTRiNyJ9.Nn5xtJd5OG9xeJ8F1iiToPeKzaxwty5jj5nTg8fzQXjn0BtxUhuvJZ1L2XGEKW6fj2g5iNZ1LfWOmXsvrqak5g

Competing Interest

The authors declare no competing interests.

References

[1] R. Nagar *et al.*, "Recent developments in state-of-the-art hydrogen energy technologies – Review of hydrogen storage materials," *Sol. Compass*, vol. 5, p. 100033, Mar. 2023, doi: 10.1016/j.solcom.2023.100033.

[2] S.-Y. Lee, J.-H. Lee, Y.-H. Kim, J.-W. Kim, K.-J. Lee, and S.-J. Park, "Recent Progress Using Solid-State Materials for Hydrogen Storage: A Short Review," *Processes*, vol. 10, no. 2, p. 304, Feb. 2022, doi: 10.3390/pr10020304.

[3] A. R. Yuvaraj, A. Jayarama, D. Sharma, S. S. Nagarkar, S. P. Duttagupta, and R. Pinto, "Role of metal-organic framework in hydrogen gas storage: A critical review," *Int. J. Hydrog. Energy*, vol. 59, pp. 1434–1458, Mar. 2024, doi: 10.1016/j.ijhydene.2024.02.060.

[4] V. A. Yartys and M. V. Lototskyy, "Laves type intermetallic compounds as hydrogen storage materials: A review," *J. Alloys Compd.*, vol. 916, p. 165219, Sep. 2022, doi: 10.1016/j.jallcom.2022.165219.

[5] K. S. Nivedhitha *et al.*, "Advances in hydrogen storage with metal hydrides: Mechanisms, materials, and challenges," *Int. J. Hydrog. Energy*, vol. 61, pp. 1259–1273, Apr. 2024, doi: 10.1016/j.ijhydene.2024.02.335.

[6] J. Gebhardt, F. Viñes, P. Bleiziffer, W. Hieringer, and A. Görling, "Hydrogen storage on metal oxide model clusters using density-functional methods and reliable van der Waals corrections," *Phys. Chem. Chem. Phys.*, vol. 16, no. 11, p. 5382, 2014, doi: 10.1039/c3cp54704c.

[7] R. B. Goncalves, R. Q. Snurr, and J. T. Hupp, "Computational Investigation of Metal Oxides as Candidate Hydrogen Storage Materials," *J. Phys. Chem. C*, vol. 126, no. 44, pp. 18661–18669, Nov. 2022, doi: 10.1021/acs.jpcc.2c04556.

[8] S. M. A.-A. Ibrahim, "Hydrogen storage in proton-conductive perovskite-type oxides and their application," *Korean J. Chem. Eng.*, vol. 31, no. 10, pp. 1792–1797, Oct. 2014, doi: 10.1007/s11814-014-0081-8.

[9] A. Salehabadi, M. I. Ahmad, N. Morad, M. Salavati-Niasari, and M. Enhessari, "Electrochemical hydrogen storage properties of Ce0.75Zr0.25O2 nanopowders synthesized by sol-gel method," *J. Alloys Compd.*, vol. 790, pp. 884–890, Jun. 2019, doi: 10.1016/j.jallcom.2019.03.160.

[10] S. A. Lachini and A. Eslami, "Excellent electrochemical hydrogen storage capabilities of green synthesized LiCoO2 nanoparticles," *Int. J. Hydrog. Energy*, vol. 51, pp. 51–60, Jan. 2024, doi: 10.1016/j.ijhydene.2023.10.286.

[11] J. Henao, Y. Pacheco, O. Sotelo, M. Casales, and L. Martinez-Gómez, "Lanthanum titanate nanometric powder potentially for rechargeable Ni-batteries: Synthesis and electrochemical hydrogen storage," *J. Mater. Res. Technol.*, vol. 8, no. 1, pp. 759–765, Jan. 2019, doi: 10.1016/j.jmrt.2018.05.019.

[12] S. Zinatloo-Ajabshir, M. S. Morassaei, O. Amiri, M. Salavati-Niasari, and L. K. Foong, "Nd2Sn2O7 nanostructures: Green synthesis and characterization using date palm extract, a potential electrochemical hydrogen storage material," *Ceram. Int.*, vol. 46, no. 11, pp. 17186–17196, Aug. 2020, doi: 10.1016/j.ceramint.2020.03.014.

[13] S. Zinatloo-Ajabshir, M. S. Morassaei, and M. Salavati-Niasari, "Simple approach for the synthesis of Dy2Sn2O7 nanostructures as a hydrogen storage material from banana juice," *J. Clean. Prod.*, vol. 222, pp. 103–110, Jun. 2019, doi: 10.1016/j.jclepro.2019.03.023.

[14] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Phys. Rev.*, vol. 136, no. 3B, pp. B864–B871, Nov. 1964, doi: 10.1103/PhysRev.136.B864.

[15] W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Phys. Rev.*, vol. 140, no. 4A, pp. A1133–A1138, Nov. 1965, doi: 10.1103/PhysRev.140.A1133.

[16] J. P. Perdew, M. Ernzerhof, and K. Burke, "Rationale for mixing exact exchange with density functional approximations," *J. Chem. Phys.*, vol. 105, no. 22, pp. 9982–9985, Dec. 1996, doi: 10.1063/1.472933.

[17] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Phys. Rev. Lett.*, vol. 77, no. 18, pp. 3865–3868, Oct. 1996, doi: 10.1103/PhysRevLett.77.3865.

[18] S. Bhattacharjee, P. Das, S. Ram, and S.-C. Lee, "A Hybrid Machine Learning Framework for Predicting Hydrogen Storage Capacities in Metal Hydrides:

Unsupervised Feature Learning with Deep Neural Networks," *ACS Appl. Mater. Interfaces*, vol. 17, no. 20, pp. 29681–29694, May 2025, doi: 10.1021/acsami.5c03612.

[19] A. Verma, N. Wilson, and K. Joshi, "Solid state hydrogen storage: Decoding the path through machine learning," *Int. J. Hydrog. Energy*, vol. 50, pp. 1518–1528, Jan. 2024, doi: 10.1016/j.ijhydene.2023.10.056.

[20] Y. Ding, L. Tong, X. Liu, Y. Liu, and Y. Zhao, "Artificial Intelligence-Driven Innovations in Hydrogen Storage Technology," *ENERGY Environ. Mater.*, vol. 8, no. 5, p. e70041, Sep. 2025, doi: 10.1002/eem2.70041.

[21] A. Goetz *et al.*, "Addressing materials' microstructure diversity using transfer learning," *Npj Comput. Mater.*, vol. 8, no. 1, p. 27, Feb. 2022, doi: 10.1038/s41524-022-00703-z.

[22] J. Hu, D. Liu, N. Fu, and R. Dong, "Realistic material property prediction using domain adaptation based machine learning," *Digit. Discov.*, vol. 3, no. 2, pp. 300–312, 2024, doi: 10.1039/D3DD00162H.

[23] D. D. Gupta *et al.*, "Leveraging Domain Adaptation for Accurate Machine Learning Predictions of New Halide Perovskites," Jan. 19, 2024, *arXiv*: arXiv:2401.10998. doi: 10.48550/arXiv.2401.10998.

[24] M. Witman, M. Allendorf, and V. Stavila, "Database for machine learning of hydrogen storage materials properties." Zenodo, Nov. 15, 2022. doi: 10.5281/ZENODO.7324809.

[25] M. Witman *et al.*, "Extracting an Empirical Intermetallic Hydride Design Principle from Limited Data via Interpretable Machine Learning," *J. Phys. Chem. Lett.*, vol. 11, no. 1, pp. 40–47, Jan. 2020, doi: 10.1021/acs.jpclett.9b02971.

[26] A. Jain *et al.*, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Mater.*, vol. 1, no. 1, p. 011002, Jul. 2013, doi: 10.1063/1.4812323.

[27] A. A. Emery and C. Wolverton, "High-throughput DFT calculations of formation energy, stability and oxygen vacancy formation energy of ABO3 perovskites," *Sci. Data*, vol. 4, no. 1, p. 170153, Oct. 2017, doi: 10.1038/sdata.2017.153.

[28] Y. Li *et al.*, "Center-Environment Feature Model for Machine Learning Study of Spinel Oxides Based on First-Principles Computations," *J. Phys. Chem. C*, vol. 124, no. 52, pp. 28458–28468, Dec. 2020, doi: 10.1021/acs.jpcc.0c06958.

[29] B. Baldassarri *et al.*, “Oxygen Vacancy Formation Energy in Metal Oxides: High Throughput Computational Studies and Machine Learning Predictions,” Sep. 03, 2023, *arXiv*: arXiv:2309.01160. doi: 10.48550/arXiv.2309.01160.

[30] L. Ward, “A general-purpose machine learning framework for predicting,” *Npj Comput. Mater.*, 2016.

[31] S. Kauwe, Andrew Falkowski, and A. Wang, “Composition-Based Feature Vector Python Package.” 2021. [Online]. Available: https://pypi.org/project/CBFV/

[32] A. de Mathelin *et al.*, “ADAPT : Awesome Domain Adaptation Python Toolbox,” 2021, *arXiv*. doi: 10.48550/ARXIV.2107.03049.

[33] M. Sugiyama, S. Nakajima, H. Kashima, P. V. Buenau, and M. Kawanabe, “Direct Importance Estimation with Model Selection and Its Application to Covariate Shift Adaptation”.

[34] T. Kanamori, I. N.-U. A. Jp, S. Hido, M. Sugiyama, and C. T. A. Jp, “A Least-squares Approach to Direct Importance Estimation”.

[35] M. Loog, “Nearest Neighbor-based Importance Weighting,” Feb. 03, 2021, *arXiv*: arXiv:2102.02291. doi: 10.48550/arXiv.2102.02291.

[36] L. Ouyang and A. Key, “Maximum Mean Discrepancy for Generalization in the Presence of Distribution and Missingness Shift,” Mar. 01, 2022, *arXiv*: arXiv:2111.10344. doi: 10.48550/arXiv.2111.10344.

[37] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, “Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD),” *JOM*, vol. 65, no. 11, pp. 1501–1509, Nov. 2013, doi: 10.1007/s11837-013-0755-4.

[38] T. Liu, N. P. Holzapfel, and P. M. Woodward, “Understanding structural distortions in hybrid layered perovskites with the $n = 1$ Ruddlesden–Popper structure,” *IUCrJ*, vol. 10, no. 4, pp. 385–396, Jul. 2023, doi: 10.1107/S2052252523003743.

[39] L. Sun *et al.*, “Perovskite-Type $CaVO_3$ Nanocomposite as High-Performance Anode Material for Lithium-Ion Batteries,” *Nano Lett.*, vol. 24, no. 49, pp. 15525–15532, Dec. 2024, doi: 10.1021/acs.nanolett.4c03328.

[40] P. Giannozzi *et al.*, “Advanced capabilities for materials modelling with Quantum ESPRESSO,” *J. Phys. Condens. Matter*, vol. 29, no. 46, p. 465901, Nov. 2017, doi: 10.1088/1361-648X/aa8f79.

[41] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, “A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu,” *J. Chem. Phys.*, vol. 132, no. 15, p. 154104, Apr. 2010, doi: 10.1063/1.3382344.

[42] K. F. Garrity, J. W. Bennett, K. M. Rabe, and D. Vanderbilt, “Pseudopotentials for high-throughput DFT calculations,” *Comput. Mater. Sci.*, vol. 81, pp. 446–452, Jan. 2014, doi: 10.1016/j.commatsci.2013.08.053.

[43] L. Yan *et al.*, “Cation ordering within the perovskite block of a six-layer Ruddlesden-Popper oxide from layer-by-layer growth – artificial interfaces in complex unit cells,” *Chem Sci*, vol. 2, no. 2, pp. 261–272, 2011, doi: 10.1039/C0SC00482K.

[44] G. Henkelman, A. Arnaldsson, and H. Jónsson, “A fast and robust algorithm for Bader decomposition of charge density,” *Comput. Mater. Sci.*, vol. 36, no. 3, pp. 354–360, Jun. 2006, doi: 10.1016/j.commatsci.2005.04.010.

[45] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[46] C. Zlotea, A. Bouzidi, J. Montero, G. Ek, and M. Sahlberg, “Compositional effects on the hydrogen storage properties in a series of refractory high entropy alloys,” *Front. Energy Res.*, vol. 10, p. 991447, Oct. 2022, doi: 10.3389/fenrg.2022.991447.

[47] V. Korostelev, J. Wagner, and K. Klyukin, “Simple local environment descriptors for accurate prediction of hydrogen absorption and migration in metal alloys,” Aug. 16, 2023, *Chemistry*. doi: 10.26434/chemrxiv-2023-x6466.

[48] A. A. Qahtan, B. Alharbi, S. Wang, and X. Zhang, “A PCA-Based Change Detection Framework for Multidimensional Data Streams: Change Detection in Multidimensional Data Streams,” in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney NSW Australia: ACM, Aug. 2015, pp. 935–944. doi: 10.1145/2783258.2783359.

[49] I. Goldenberg and G. I. Webb, “PCA-based drift and shift quantification framework for multidimensional data,” *Knowl. Inf. Syst.*, vol. 62, no. 7, pp. 2835–2854, Jul. 2020, doi: 10.1007/s10115-020-01438-3.

[50] E. M. Mirkes, J. Bac, A. Fouché, S. V. Stasenko, A. Zinovyev, and A. N. Gorban, “Domain Adaptation Principal Component Analysis: Base Linear Method for Learning

with Out-of-Distribution Data," *Entropy*, vol. 25, no. 1, p. 33, Dec. 2022, doi: 10.3390/e25010033.

[51] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised Visual Domain Adaptation Using Subspace Alignment," in *2013 IEEE International Conference on Computer Vision*, Sydney, Australia: IEEE, Dec. 2013, pp. 2960–2967. doi: 10.1109/ICCV.2013.368.

[52] A. Gencer, G. Surucu, and S. Al, "MgTiO3Hx and CaTiO3Hx perovskite compounds for hydrogen storage applications," *Int. J. Hydrog. Energy*, vol. 44, no. 23, pp. 11930–11938, May 2019, doi: 10.1016/j.ijhydene.2019.03.116.

[53] Q. Zhu, W. Huang, C. Huang, L. Gao, Y. Su, and L. Qiao, "The d band center as an indicator for the hydrogen solution and diffusion behaviors in transition metals," *Int. J. Hydrog. Energy*, vol. 47, no. 90, pp. 38445–38454, Nov. 2022, doi: 10.1016/j.ijhydene.2022.09.021.

[54] A. Gencer and G. Surucu, "Properties of BaYO3 perovskite and hydrogen storage properties of BaYO3Hx," *Int. J. Hydrog. Energy*, vol. 45, no. 17, pp. 10507–10515, Mar. 2020, doi: 10.1016/j.ijhydene.2019.06.198.

[55] F. S. Sangsefidi and M. Salavati-Niasari, "Fe_2O_3 –CeO_2 Ceramic Nanocomposite Oxide: Characterization and Investigation of the Effect of Morphology on Its Electrochemical Hydrogen Storage Capacity," *ACS Appl. Energy Mater.*, vol. 1, no. 9, pp. 4840–4848, Sep. 2018, doi: 10.1021/acsaem.8b00907.

[56] Y. Chen, J. Fan, T. Liu, J. Wang, M. Zhang, and C. Zhang, "Theoretical study on the effect of an O vacancy on the hydrogen storage properties of the LaFeO3 (010) surface," *Int. J. Hydrog. Energy*, vol. 44, no. 11, pp. 5374–5381, Feb. 2019, doi: 10.1016/j.ijhydene.2018.09.097.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFiles.docx