Prediction of the optimal hydrogen storage in TiVCrMnAlFe alloys using machine learning models

Ashley Phala [1*,2,3], *David* Tshwane [2], *Charles* Siyasiya[1], *Sisa* Pityana[3] and *Regina* Maphanga[2]

[1]Department of Material Science and Metallurgical Engineering, University of Pretoria, South Africa
[2]Next Generation Enterprise and Institutions, Council for Scientific and Industrial Research, 0001, South Africa,
[3]CSIR, Photonics Centre, Laser Enabled Manufacturing, Pretoria Campus, South Africa

Abstract. The quest for efficient hydrogen storage materials is vital for the advancement of clean energy technologies. Among various candidates, high entropy alloys (HEAs) such as TiVCrFeAl have received significant attention due to their tunable structures and favorable thermodynamic properties. This study explores the application of machine learning (ML) techniques to predict the hydrogen storage capacity of the TiVCrFeAl alloy system. Using a dataset compiled HEAPS and calculated properties, multiple regression models, such as Random Forest, Gradient Boosting, Decision Tree, XGBoost, Linear Regression and Support Vector Regression, were trained to capture complex relationships between alloy composition, processing parameters, and hydrogen weight percent (wt%). Data preprocessing steps included feature selection, imputation of missing values, and standardization to ensure model robustness. The performance of the models was evaluated using cross-validation and test set metrics such as R^2 and mean squared error. Results show that ensemble-based models, particularly Random Forest and XGBoost, achieved high predictive accuracy, demonstrating the effectiveness of ML in modeling nonlinear property trends in HEAs. This approach offers a powerful tool for screening and optimizing hydrogen storage materials, accelerating the discovery process through computational insight.

1 Introduction

The recent worldwide shift toward decarbonization and sustainable energy systems has brought hydrogen into focus as a clean, high-energy-density fuel. As the world transitions from fossil fuels, hydrogen's versatility across a variety of applications, such as transportation, industrial processes, and grid-scale energy storage, positions it as a key player in achieving carbon neutrality [1,2]. However, solving some of the most important challenges facing hydrogen technologies, such as safety, high storage capacity, and effectiveness, is essential to their broad acceptance. Conventional storage methods, like high-pressure gas

*Corresponding Author: mohlagashanephala@gmail.com

tanks and cryogenic liquid storage, remain predominant but are accompanied by significant safety and technical concerns, which include high energy consumption for compression and liquefication, leakage risks, as well as material degradation [3,4].

Metal hydride-based solid-state hydrogen storage offers a promising solution with reduced operating pressures, higher volumetric densities, and intrinsic safety. Metal hydrides integrate hydrogen atoms into their interstitial sites, which enables compact and reversible hydrogen storage, while mitigating operational hazards and energy losses [5]. Transitional metal hydrides, such as Ti-V-Cr systems, among various classes of hydrogen storage materials, have gained significant attention. However, they face several setbacks that limit their large-scale application. These alloys often require high operating temperatures (above 100 °C) and high hydrogen pressures to function efficiently, which increases energy consumption and system complexity. The cost of raw materials, particularly vanadium, is high, making large-scale use expensive. Although their volumetric capacity is good, their gravimetric hydrogen density is relatively low compared to lighter materials such as magnesium hydrides, limiting their suitability for mobile applications. They also typically require activation treatments, such as repeated hydrogenation–dehydrogenation cycles, before reaching optimal performance. Furthermore, Ti–V–Cr hydrides are sensitive to impurities like oxygen, carbon monoxide, and moisture, which can poison the surface and degrade performance

These alloys are known for forming body-centred cubic (BCC) solid state solutions, which promote fast hydrogen desorption/absorption kinetics and high hydrogen storage capacities [6,7]. Nonetheless, their performance must be improved further to meet practical requirements. Alloying with additional elements such as Mn, Fe, and Al has proven to effectively improve thermodynamic, structural properties, and stabilizes the BCC phase, which improves the hydrogen cycling stability [8-10]. Al and Mn help to lower the plateau pressure, which improves the phase stability, whereas Fe improves the mechanical properties. However, the enormous compositional design space created by multicomponent alloys imposes a major challenge in the traditional experimental optimisation. Certain key parameters, such as mixing entropy (ΔS_{mix}), valence electron concentration (VEC) and average atomic size mismatch (δ), play a crucial role in determining the hydrogen storage performance and phase formation of multicomponent alloys. These parameters allow rapid prediction of key alloy properties and are vital for data-driven approaches for material discovery.

In such a large compositional space, traditional trial-and-error approaches become limited. As such, Machine learning (ML) techniques have emerged as a transformative tool in material science, offering ways to discover new functional materials with tailored properties, and optimise compositions with significantly fewer experiments [11,12]. In the context of hydrogen storage research, ML has been used in various forms, such as for the prediction of thermodynamic properties (i.e. hydride formation enthalpies) [13], estimating hydrogen storage capacities [14], and to classify hydride-forming alloys [15]. Nonetheless, most recent work focuses on simple binary or ternary systems, due to challenges such as sparse datasets and feature complexity. ML applications to high-entropy alloys or multicomponent systems have been rare. Zhang *et al*. [16] have applied random forest regression to predict the hydrogen storage capacities in high entropy alloys by using features such as atomic radius, enthalpy of formation, among others. Xie *et al*. [17] have deployed deep learning algorithms to classify BCC forming high entropy alloys for hydrogen storage applications, but no experimental validation was done. Regardless of these advances, no systemic ML study has been conducted to target the TiVCrMnAlFe alloys system, a compositional complex with highly promising hydrogen storage capacity, given the favourable hydrogen storage capacity of Ti-V-Cr and stabilizing effects of Mn, Fe, and Al.

In this study, ML framework was developed to predict hydrogen storage capacities in TiVCrMnAlFe alloys, the equiatomic TiVCrMnAlFe alloy has a mixing entropy of about 1.79 R, which is greater than 1.5 R. Therefore, it is classified as a high-entropy alloy. By constructing a curated datasets using High-Entropy Alloy Prediction Software) (HEAPS) [18] and calculating compositional features such as δ, ΔS_{mix}, VEC and electronegativity differences, we train and validate regression models to identify optimal compositions with high hydrogen storage capacities. Our approach not only uncovers underlying composition-property relationships but also proposes new alloy formulations with improved hydrogen storage potential. Through this work, we aim to contribute to the rapid development of high-performance hydrogen storage materials essential for realizing sustainable, hydrogen-based energy systems.

2 Methodology

In this study, machine learning (ML) techniques were used to predict the hydrogen storage capacities of TiVCrMnAlFe alloys. A dataset of 983 TiVCrMnAlFe alloys were generated using HEAPS, which simulated thermodynamic properties, phase stability, and the pressure-composition-temperature (PCT) curve was used to estimate the hydrogen storage capacities at standard conditions. Critical physical and chemical descriptors such as the average atomic size mismatch (δ), valence electron concentration (VEC), electronegativity difference (Δχ), mixing entropy (ΔS_{mix}), and estimated enthalpy of mixing (ΔH_{mix}), were extracted for each alloy. These featured were selected due to their crucial role in hydrogen storage, phase stability and diffusion kinetics. Various supervised ML algorithms such as Random Forest Regression, Decision Tree, Linear Regression, Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR) and XGBoost were developed. The dataset was split randomly into training and testing subsets with an 80/20 ratio, and the model robustness was ensured through five-fold cross-validation. Prediction accuracy was enhanced by optimizing hyperparameters via grid randomized search methods, coefficient of determination (R^2) and mean absolute error (MAE) were used to evaluate the model performance. Feature importance analysis was performed for the tree-based models to identify key descriptors influencing hydrogen storage capacity. Figure 1 below show a ML workflow for hydrogen storage prediction.

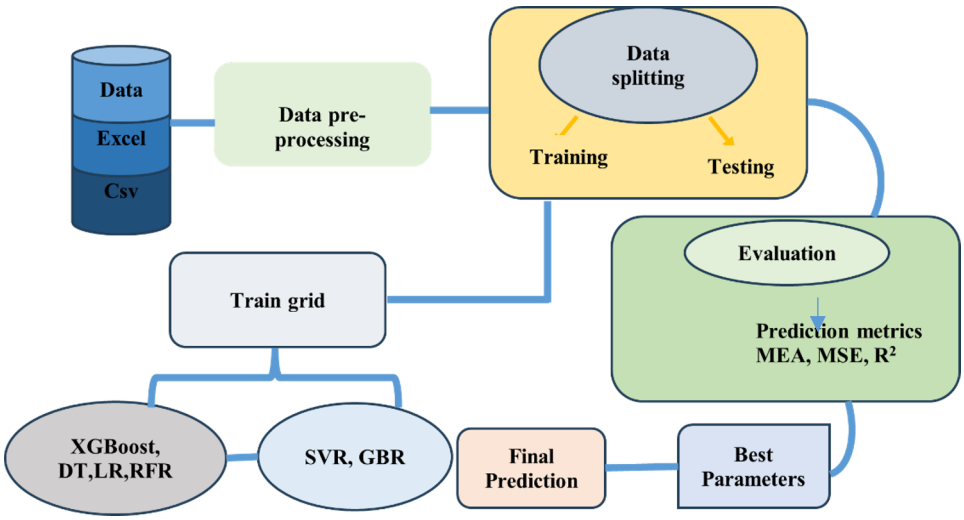

Fig. 1. Machine learning workflow framework.

2.1 Cross-validation

In ML cross-validation is used to improve the performance of the models and mitigates overfitting by dividing the dataset into multiple subsets. In each iteration the model is trained on a portion of data (the training set) and evaluated on the remaining portion (the testing set), which ensures that the model's predictive performance is rigorously assessed across the different subsets [19]. To improve the evaluation, the splitting process is repeated systematically, and different subsets are selected during each iteration [20]. In particular, the K-fold cross-validation reduces bais from random data splitting and achieved an optimal balance between variance and computational efficiency [21]. In K-fold validation, the dataset is separated into k equal parts, in each iteration one fold serves as the testing set, whereas the remaining k -l folds are used as the training set. Commonly used k is 5 and 10, setting k=n corresponds to leave-one-out cross-validation [22].

In this study, 5-fold cross-validation was used to minimize overfitting when predicting the hydrogen storage properties, as shown in Figure. 2. Four subsets were used for training and the remaining subset was used for testing in each fold, to ensure that each subset served as the test set this procedure was repeated 5 times. The final model accuracy was determined by averaging the results across all five folds. This iterative process not only tunes hyperparameters effectively but also results in a more accurate and robust model for predicting both the valence electron concentration (VEC) and hydrogen storage capacities (h-wt%). By systematically optimizing hyperparameters across multiple validation folds, cross-validation ensures minimal risk of overfitting and enhances the model's generalization ability.

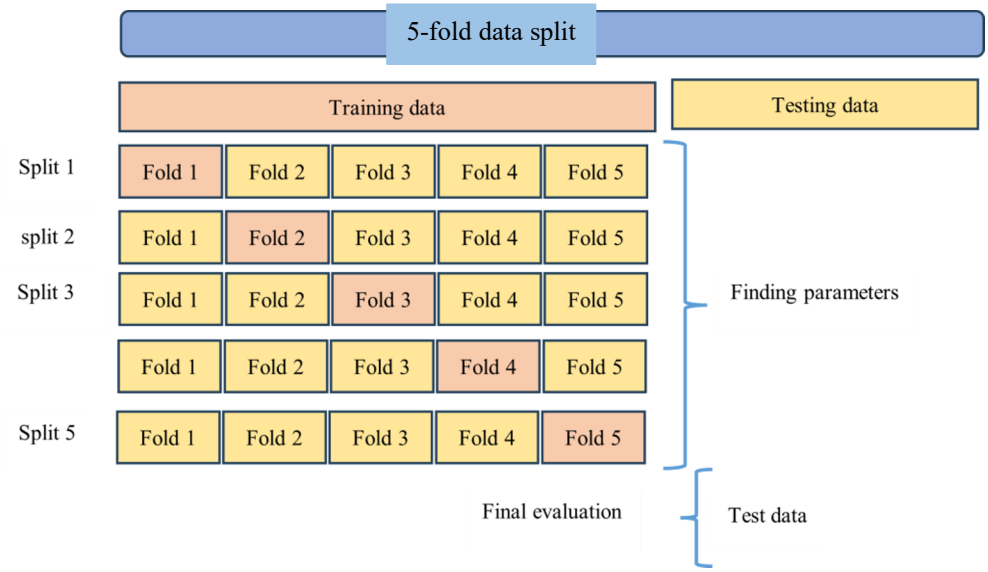

Fig. 2. Cross-validation 5-fold data split.

3 Results and discussion

3.1 Data analyses

Figure 3 demonstrate the distribution of hydrogen storage capacity (wt%) in TiVCrMnAlFe HEAs. The results show a right-skewed distribution with most materials exhibiting

absorption capacities between 4.0 and 4.2 wt%. The peak absorption is centered around 4.1 wt%, suggesting that a significant proportion of the dataset comprises materials with relatively high hydrogen storage potential. Very few materials demonstrate absorption below 3.8 wt%, indicating that low-capacity materials are less prevalent in the dataset. The shape of the distribution suggests that the dataset is biased toward higher-performing compositions. This trend supports the potential of these materials as strong candidates for hydrogen storage applications.

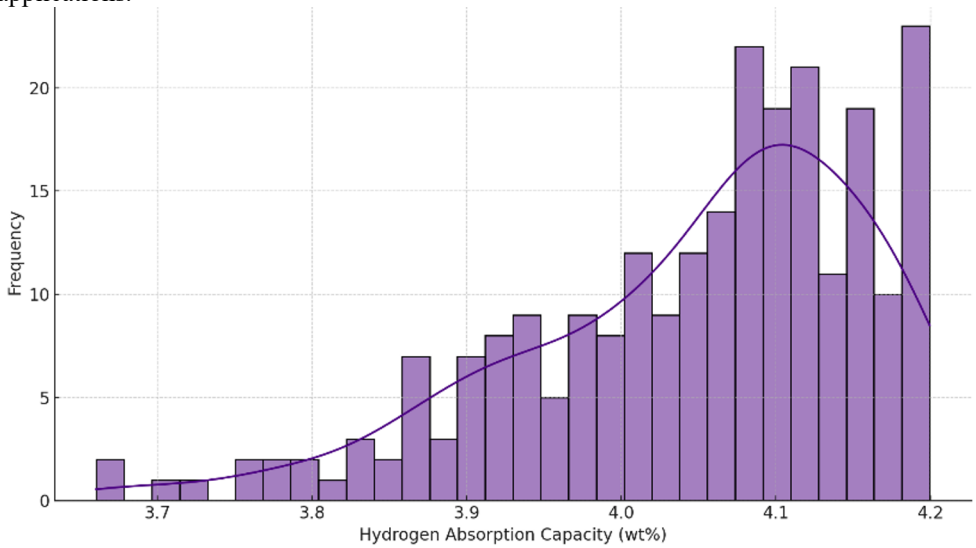

Fig. 3. Shows the distribution of hydrogen storage in TiVCrMnAlFe alloys.

3.2 Feature importance

Several fundamental material descriptors for hydrogen storage capacity in HEAs such as δ (Atomic Size Mismatch), VEC (Valence Electron Concentration, Δχ (Electronegativity Difference), ΔS_{mix} (Mixing Entropy) and ΔH_{mix} (Mixing Enthalpy), are used as shown in figure 4. The feature selection is determined by the data available and the impact of the parameter on the target [23,24]. These parameters are important for understanding the phase formation and stability of HEAs and for predicting different properties such as the hydrogen storage capacities. The compositional formulas are also considered for training and prediction, VEC and the hydrogen storage capacities are the targets.

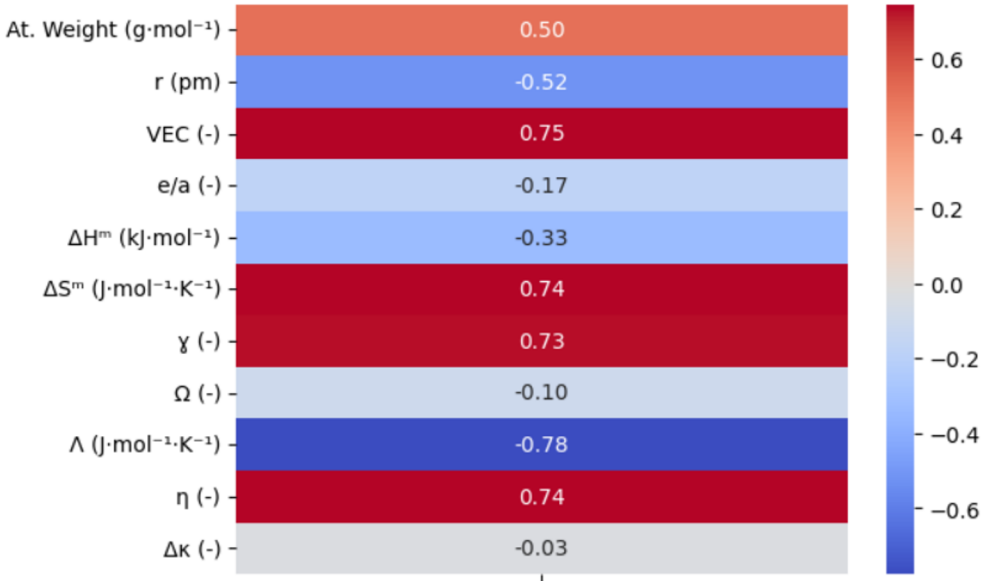

Fig. 4. Present the feature importance in high entropy alloys for hydrogen storage capacity predication.

3.3 H_2 wt% prediction

Table 1 shows the results of the model evaluation demonstrating strong predictive performance across all the tested algorithms, with Gradient Boosting (GB) and XGBoost emerging as the top performers. XGBoost achieved the highest accuracy, with a cross-validation R^2 (CV R^2) of 0.998 and a minimal mean squared error (MSE) of 0.0003, indicating highly precise predictions. Gradient Boosting followed closely with a CV R^2 of 0.997 and a slightly lower MSE of 0.0002, showcasing excellent performance with well-optimized parameters (learning_rate = 0.1, max_depth = 5, n_estimators = 100). Random Forest Regressor (RFR) also performed impressively, recording a CV R^2 of 0.997 and MSE of 0.0003 using max_depth = 10 and n_estimators = 200. Decision Tree (DT), while slightly less accurate with a CV R^2 of 0.985 and MSE of 0.0014, still achieved a strong R^2 of 0.988 on the test set without limiting tree depth. Support Vector Regression (SVR) produced the lowest performance among the models, with a CV R^2 of 0.967 and MSE of 0.0038, though it remained effective considering its simplicity. Overall, ensemble methods such as GB, XGBoost, and RFR demonstrated superior capability in modeling the data accurately, particularly when optimized with appropriate hyperparameters.

Table 1. shows the metrices for different algorithms.

Models	CV R^2	MSE	R^2	Best Parameters
RFR	0.997	0.0003	0.997	'Max_depth' :10 'n_esimator' :200
DT	0.985	0.0014	0.988	'Max_depth' : None
GB	0.997	0.0002	0.998	'Learning_rate' : 0.1 'Max_depth' : 5 'n_esimator' : 100

XGBoost	0.998	0.0003	0.998	'Learning_rate' : 0.1 'Max_depth' : 5 'n_esimator' : 100
SVR	0.967	0.0038	0.967	'C':1 'epsilon' : 0.1

The "Predicted vs Actual" graph visually compares the performance of five machine learning models, Random Forest Regressor (RFR), Decision Tree (DT), Gradient Boosting (GB), XGBoost, and Support Vector Regressor (SVR as shown in figure 5 below. Each model's predictions are represented using distinct bright-colored dots, making it easy to differentiate between them. The dotted black line represents the ideal scenario where predicted values exactly match the actual values (i.e., Predicted = Actual). Models with points that closely follow this line are performing well.

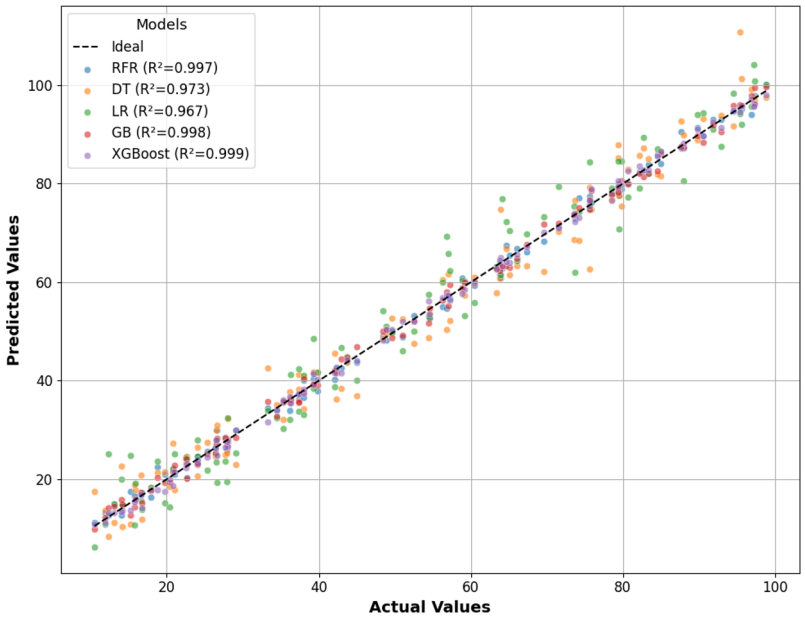

Fig. 5. The graph illustrates visually comparison (R^2) between five machine learning models.

3.4 VEC predictions

The table 2 presents a comparison of five machine learning models Random Forest Regressor (RFR), Decision Tree (DT), Linear Regression (LR), Gradient Boosting (GB), and XGBoost evaluated on their ability to predict a target variable. The performance metrics used include Cross-Validation R^2 (CV R^2), Mean Squared Error (MSE), and R^2 on the test set, along with the best hyperparameters for each mode. XGBoost shows the lowest MSE (0.0163) and a very high R^2 score (0.994), indicating that it has the best overall performance. Its hyperparameters moderate learning rate, max depth of 5, and 100 estimators suggest a well-regularized, deep model capable of capturing complex patterns without overfitting. Random Forest Regressor (RFR) follows closely with an excellent R^2 score of 0.996 and a slightly higher MSE of 0.0264. It also has the highest cross-validation R^2 (0.979), suggesting consistent performance across different subsets of the data. Its ensemble nature and use of 200 estimators help balance bias and variance effectively.

Gradient Boosting (GB) performs very well with an R^2 of 0.991 and MSE of 0.0253, comparable to RFR. The model uses a learning rate of 0.1 and a shallower tree (max depth 3), suggesting a model that builds upon weak learners in a stable and robust way. Decision Tree (DT) has a decent R^2 score of 0.978, but a higher MSE of 0.0622 compared to the ensemble methods. The absence of specific parameters implies it might be using default settings. While simpler and faster, DTs can be prone to overfitting or underfitting depending on depth and pruning settings. Linear Regression (LR) has the lowest test R^2 (0.962) and the highest MSE (0.1075) among all models. Despite this, its cross-validation R^2 is still competitive (0.975). This performance shows that while LR can capture general trends, it struggles to model non-linear or complex relationships in the data, even with regularization (C=10) and a small epsilon for error tolerance.

Table 2. shows the metrices for different algorithms.

Algorithms	CV R^2	MSE	R^2	BEST PARAMETERS
RFR	0.979	0.0264	0.996	'n_esimators': 200
DT	0.948	0.0622	0.978	None
LR	0.975	0.1075	0.962	'C':10 'epsilon':0.1
GB	0.977	0.0253	0.991	'Learning_rate': 0.1 'Max_depth': 3 'n_esimators': 100
XGBoost	0.977	0.0163	0.994	'Learning_rate': 0.1 'Max_depth': 5 'n_estimators': 100

The "Predicted vs Actual" graph visually compares the performance of five machine learning models, Random Forest Regressor (RFR), Decision Tree (DT), Gradient Boosting (GB), XGBoost, and Linear Regression as shown in figure 6 below. Each model's predictions are represented using distinct bright-colored dots, making it easy to differentiate between them. The dotted black line represents the ideal scenario where predicted values exactly match the actual values (i.e., Predicted = Actual). Models with points that closely follow this line are performing well.

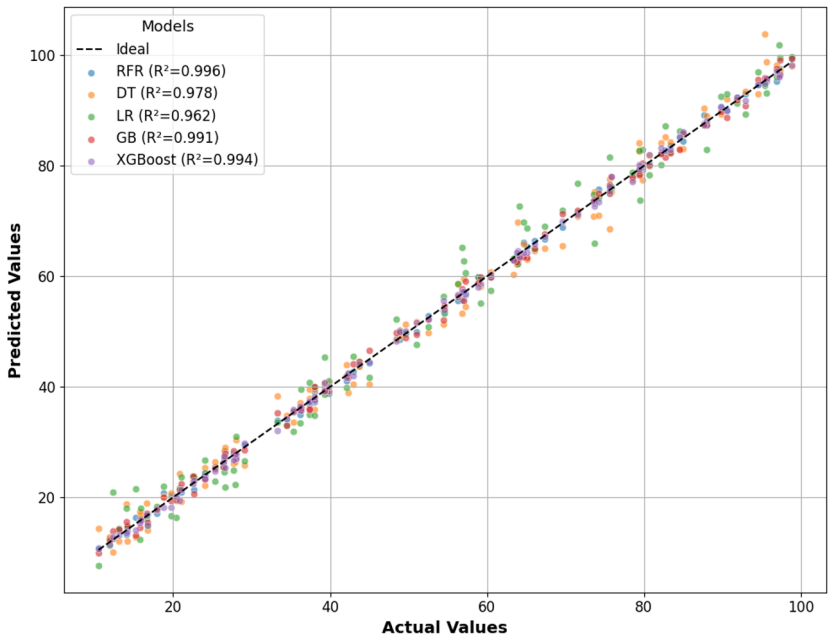

Fig. 6. Present the **p**redicted vs actual" graph visually compares the performance of five machine learning models, Random Forest Regressor (RFR), Decision Tree (DT), Gradient Boosting (GB), XGBoost, and Linear Regression.

4 Conclusion

Machine learning models, when trained on curated datasets of HEA compositions and properties, can accurately predict the hydrogen storage performance of complex alloys like TiVCrFeAl. Among the models tested, ensemble methods such as Random Forest and XGBoost showed superior predictive power, capturing the intricate interplay between elemental constituents and hydrogen uptake. The study highlights the potential of ML not only as a predictive tool but also as a strategic guide for experimental design, reducing the need for costly and time-consuming synthesis trials. This data-driven framework supports rapid materials screening, paving the way for the accelerated development of next-generation hydrogen storage solutions.

The authors would like to thank my colleagues at Laser Enabled Manufacturing (LEM) and Next Generation Enterprises at CSIR for their unwavering support and resources in conducting this study, as well as the Department of Materials Science and Metallurgical Engineering at the University of Pretoria.

Data will be made available on request.

References

1. J.A. Turner, Sustainable hydrogen production. Sci., **305** 972 (2004). DOI:10.1126/science.1103197
2. I. Staffell, D. Scamman, A. V. Abad, P Balcombe, P. Dodds, P Ekins, N Shah and K. R Ward, The role of hydrogen and fuel cells in the global energy system. Energy. Environ. Sci., **12** 463 (2019). https://doi.org/10.1039/C8EE01157E

3. L. Schlapbach and A. Züttel Hydrogen-storage materials for mobile applications. Nature, **414** 353 (2001) DOI:10.1142/9789814317665_0038

4. B. Sakintuna, F. Lamari-Darkrim, and M Hirscher, Metal hydride materials for solid hydrogen storage, A review. Int. J. Hydrog. Energy, **9** 32 (2007) DOI:10.1016/j.ijhydene.2006.11.022
5. [5] M. Lototskyy, V. A Yartys, B. G Pollet and R. C Bowman, Metal hydride hydrogen compressors, A review. Int. J. Hydrog. Energy, **11** 39 {(2004) DOI.org:10.1016/j.ijhydene.2014.01.158
6. [6] H. Wang, J. Zhang and E. Akiba, Hydrogen storage properties of Ti–V–Cr alloys with body-centered cubic structure. Journal of Alloys and Compounds, **2** 377 (2004) DOI.org/10.1016/j.ijhydene.2023.07.020
7. K. Tanaka, K. Sato and T. Ichikawa, Improvement of hydrogen storage characteristics of Ti–V–Cr alloys by alloying. J. Alloys and Compd, **404** 406 (2005) DOI.org/10.1016/j.matchemphys.2024.129132
8. X. Luo, and E. Akiba, Effects of Al substitution on Ti–V–Cr BCC solid solution hydrogen storage alloys. J. Alloys and Compd, **1** 311 (2000) DOI:10.1016/j.matchemphys.2024.129132
9. X. Guo, H. Lin, and G. Liang, Synthesis and hydrogen storage properties of Ti–V–Cr–Fe–Mn BCC alloys. J. Alloys and Compd, **2** 505 (2010) DOI:10.1016/j.ijhydene.2021.06.137
10. K. T Butler, D. W Davies, H. Cartwright, O. Isayev, and A. Walsh, Machine learning for molecular and materials science. Nature, **559** 7715 (2018) DOI: 10.1038/s41586-018-0337-2
11. T. Lookman, P. V. Balachandran, D. Xue and R. Yuan, Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. Npj Comput. Mater., **1** 5 (2019) DOI: 10.1038/s41524-019-0153-8
12. C. Kim, G. Pilania and R. Ramprasad, Machine learning assisted predictions of intrinsic dielectric breakdown strength of ABX3 perovskite. J. Phys. Chem. A, **27** 120 (2016) DOI: 10.1021/acs.jpcc.6b05068.
13. M Fernandez, V Zelenay and A Balducci, Machine learning predictions of hydride formation enthalpy for energy storag. Mater. Today Energy, **20** 10069 (2021) DOI: 10.1016/j.ensm.2023.102964
14. R. Jalem, M. Nakayama, M. Wakihara, and T. Kasuga, Machine learning accelerated screening of hydride materials for hydrogen storage**.** Comput. Mater. Sci, **183** 109795 (2020) DOI: 10.1039/C3TA13235H
15. L. Ma, Y. Zuo, T. Xie and S. P. Ong, Machine learning for electrochemical energy materials. Energy Environ Sci**, 2** 13 (2020) DOI: 10.1016/j.mtsust.2025.101163
16. Y. Zhang, T. Xie, Y. Wang, and J. Sun, Machine-learning predictions of hydrogen storage capacities of high-entropy alloys. Mater. Sci. Technol, **82** 86 (2021) DOI:10.1016/j.pmatsci.2022.101018
17. L. Xie, Y. Zhu, W. Wang and H. Chen, Deep learning for predicting phase formation in high-entropy alloys. Acta Mater, **224** 117505 (2022) Doi.10.1016/j.jmrt.2022.01.172
18. W. Xu , Y. Wang and X. Fang, Artificial neural network prediction of plateau pressures in metal hydrides. Int. J. Hydrog. Energy, **33** 46 (2021) DOI:10.1016/j.jmrt.2022.01.172

19. B. Vyas, A. Khatiashvili, L. Galati, K. Ngo, N. Gildeener-Leapman, M. Larsen and I. K. Lednev, Raman hyperspectroscopy of saliva and machine learning for Sjögren's disease diagnostics. Sci Rep, **14** 11135 (2024) DOI: 10.1038/s41598-024-59850-6

20. R.Y. Coley, Q. Liao and N. Simon,., Empirical evaluation of internal validation methods for prediction in large-scale clinical data with rare-event outcomes: a case study in suicide risk prediction, BMC Med Res Methodol, **23** 33 (2023) DOI: 10.1186/s12874-023-01844-5
21.] K. Mnich, A. Polewko-Klim, A.K. Golińska, W. Lesiński and W.R. Rudnicki, Super learning with repeated cross validation. ICDMW IEEE **629** 635 (2020) DOI:10.48550/arXiv.2003.08342