BIBECHANA
Vol. 23, No. 2, April 2026, 84-92

ISSN 2091-0762 (Print), 2382-5340 (Online)
Journal homepage: http://nepjol.info/index.php/BIBECHANA
Publisher:Dept. of Phys., Mahendra Morang A. M. Campus (Tribhuvan University)Biratnagar

Machine learning driven prediction of formation energy of $A_xM_yM'_zO_6$ oxides

Pratima Khadka, Sabita Pandey, Madhav Prasad Ghimire*
Central Department of Physics, Tribhuvan University, Kirtipur, Kathmandu, Nepal

*Corresponding authors: Email: madhav.ghimire@cdp.tu.edu.np

Abstract
Machine learning has become an effective tool in materials discovery, offering significantly lower computational requirements than traditional density functional theory calculations and experimental approaches. This work applies machine learning to predict formation energies of $A_xM_yM'_zO_6$ oxides using a dataset of 350 compounds with 28 structural, elemental, and electronic descriptors. Four regression models such as CatBoost, Gradient Boosting, Random Forest and Support Vector Regression were trained and compared to obtain the accurate values. Among them, CatBoost achieved the highest accuracy (R^2 = 0.83 and RMSE = 0.41 eV/atom), outperforming the other approaches. Feature analysis further revealed that electronegativity, ionization energy, and band gap are the dominant factors influencing the stability of $A_xM_yM'_zO_6$ oxides. These results demonstrate the potential of machine learning to provide fast and reliable prediction of formation energies and to support the design of stable oxide materials suitable for energy devices.

Keywords
Machine learning, formation energy, Random Forest, Gradient Boosting Regression, Support Vector Regression, RMSE

Article information
Manuscript received: February 2, 2026; Revised: April 25, 2026; Accepted: April 26, 2026
DOI: https://doi.org/10.3126/bibechana.v23i2.90416

1 Introduction

Machine learning is a multidisciplinary field that can automatically learn data patterns and make desired predictions in an easier and efficient way [1]. The machine learning algorithm has high robustness and effectiveness in modeling real life problems which conventional techniques cannot handle [2]. Using the available input features, machine learning algorithms can learn the irregular pattern between variables and predict unknown target variables [3]. The algorithms aim to generalize the pattern that is obtained for any set of input variables. In the training process, the model utilizes the target data to learn the relationship between the input features and the expected outcomes [4]. Machine learning has made dramatic progress in various fields. Recently, machine learning has played a key role in materials science and has transformed this field by accelerating the discovery of materials with novel properties [5]. Traditional approaches to material discovery like density functional theory (DFT) based approach and experimental methods are expensive, inefficient, and often need a lengthy research and development cycle regarding the growing needs of materials science [6]. Machine learning can significantly lower computing costs while speeding up the materials discovery process, it has become a highly successful alternative for conventional DFT calculations and tedious experimental processes [7]. By offering a fresh knowledge of the fundamental chemical or physical relationships influencing qualities of interest, machine learning has advanced beyond property prediction to the discovery and design of innovative materials for a variety of applications and materials classes [8].

The formation energy plays an important role to create novel materials with the desired characteristics. Stability of any materials depends on its formation energy, a system is more likely to be stable if its formation energy is negative. Thus, prediction of the formation energy is essential to study the various properties of any materials. Scientists typically apply first-principles calculations and also do experimental work to acquire precise formation energy. However, calculating formation energy of complex system from first principles is both costly and time consuming. Furthermore, measuring the formation energy of a large number of materials through experiments is not feasible [9]. Therefore, machine learning plays a significant role in the prediction of the formation energy of complex systems like complex oxides [10].

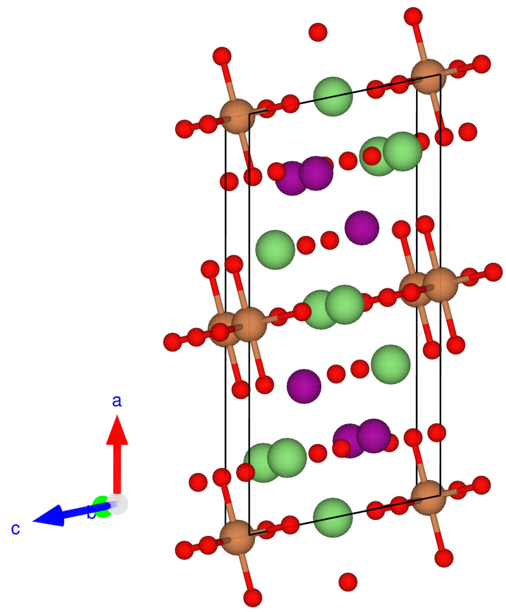

Figure 1: Crystal structure of $Li_3Fe_2SbO_6$ layered oxide.

The formation energy of a compound indicates how thermodynamically stable it is and can be calculated using the relation

$$E_f = E(A_xM_yM'_zO_6) - xE(A) - yE(M) - zE(M') - 6E(O) \quad (1)$$

where $E(A_xM_yM'_zO_6)$ represents the total energy of the oxide compound, E(A), E(M), E(M′), E(O) denote the energy per atom for the corresponding stable element A, M, M′ and O, respectively. Machine learning has been employed to predict the thermal expansion [11], thermodynamic stability of perovskite oxides [12], dielectric constant [13], band gap of ABO_3 [14], formation energy [15–18]. Pilania et al. [19] used chemostructural fingerprints in their machine learning algorithms to predict formation energy and other polymer properties. Their scatter plots showed errors of roughly 0.5 eV based on visual inspection. Ward et al. [20] introduced Matminer, which applies a random forest-based machine learning approach to estimate formation energy. Lotfi et al. [21] built an SVR model to predict the formation energy, helping researchers explore and synthesize materials in less researched composition areas. Many of these studies focus on the application of machine learning algorithms to relatively simple oxide systems and single crystal system materials.

In this study, we have used multiple regression models to predict the formation energies of a large set of $A_xM_yM'_zO_6$ layered oxides. Here, x=1-3; y,z=1,2; $y+z$=3; A are alkali metals, alkaline earth, or lanthanoids; M = 3d transition metals; and M′ being transition metals or p-block non-metals. Unlike previous studies, this work deals with a dataset with different crystal structures, complex layered oxides and systematically compares multiple models to achieve improved predictive accuracy for formation energy. Figure 1 shows the crystal structure of $Li_3Fe_2SbO_6$ layered oxide, where alternating layers of lithium ions and transition metal ions are stacked. We have explored the formation energies by applying the machine learning technique. The selected compounds are expected to have various applications including photovoltaics, superconductor, optoelectronics, energy storage, catalysis, sensor quantum materials, etc.

2 Methodology

For this work, we collected the dataset for the $A_xM_yM'_zO_6$ family of layered oxides from the materials project database [22]. We manually extracted relevant material properties from materials project to prepare the dataset used in this work. The dataset offers structural diversity for model development by including both perovskite and non-perovskite oxide structure, theoretically predicted and experimentally observed compounds.

350 $A_xM_yM'_zO_6$ layered oxide compounds with parameters such as molecular weight, atomic radii, atomic mass, formation energy, electronegativity, oxidation state, lattice angle (alpha, beta and gamma), total magnetization, band gap, energy above hall, melting point, boiling point, density and number of atoms had taken into consideration. The dataset was examined using correlation plots, boxplots and histogram to understand feature distribution and relationships before the model was trained. In order to maintain the distinctive characteristics of the materials, all data points were kept. To train machine algorithm, all these features are taken as input feature except formation energy. No additional feature extraction or engineering techniques were applied. All features were directly obtained from the materials project database.

In this study, 12 different machine learning models were initially considered and evaluated. Based on the greater performance score only four models highlighted in the discussion. We used the Scikit-learn Python library to create a number of regression models, such as Random Forest, Gradient Boosting, Support Vector Regression, and CatBoost. A wide range of tools for data preprocessing, model construction, assessment, and implementation in machine learning applications are provided by this package. The train_test_split function was used to divide the dataset into training and testing sets. In order to increase model stability and speed, feature scaling was done using StandardScaler prior to feeding the data into the models. Hyperparameter tuning was used to reduce overfitting and enhance model performance. Hyperparameter tuning was performed using GridSearchCV with 5-fold cross validation. The standard k-fold strategy with random splitting was applied. GridSearchCV was utilized to evaluate various parameter combinations and determine the best ones. We first created a grid of possible discrete hyperparameter values, and then fitted the model using every possible combination. After logging the model performance for each set, we select the combination that yields the best outcomes [23]. R^2 and RMSE, are calculated to determine the accuracy of different models. The R^2 measure indicates how well the set of independent variables as a whole explains the target variable. The value of R^2 falls between 0 and 1. For the better prediction the value of R^2 should be near to 1. RMSE measures the square root of the average of the squared discrepancies between the prediction and the actual observation.

The coefficient of determination is given by [24]

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{2}$$

The formula for Root Mean Squared Error (RMSE) is given by [24]

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

where y_i represent the actual values, $\bar{y}$ is the corresponding average, $\hat{y}_i$ are the predicted values, and n is the number of external sets [24].

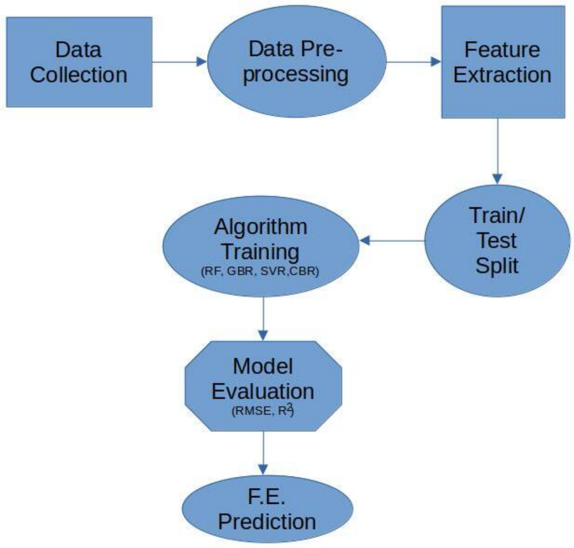

Figure 2: Flow chart of machine learning algorithm for predicting formation energy of $A_xM_yM'_zO_6$ oxides.

The step-by-step procedure adopted in this work is summarized in Figure 2. This workflow provides a systematic and reproducible approach for developing and evaluating machine learning models for formation energy prediction.

2.1 Random Forest Regressor:

Random Forest Regressor (RFR) is an ensemble learning method where multiple decision trees are combined to improve predictive performance. Instead of relying on a single model, RF constructs a collection of decision trees and aggregates their output to produce a final prediction. So, the large number of trees are especially significant in RF regression [25]. Two factors contribute to the randomness of a random forest. First, each tree is trained on a random subset of the training data. Second, each split in a tree is generated using a random subset of input features. The trees are highly unstable, the unpredictability in them causes variations in each tree's prediction. As each individual tree produces a multidimensional step function, RF stabilizes the output and provides more reliable results by averaging predictions from the individual trees.

For an input vector x, the final prediction of RF model constructs K regression trees and averages the outcomes as shown in below [26]

$$\hat{f}^K_{rf}(x) = \frac{1}{K}\sum_{K=1}^{K} T_k(x) \tag{4}$$

where $T_k(x)$ represents the prediction from the k$^{\text{th}}$ decision tree.

2.2 Gradient Boosting Regressor:

Gradient Boosting Regressor (GBR) is a powerful ensemble learning technique which allows to integrate predictions from many learner models and develop a final model with accurate prediction. The training data is carefully resample using the boosting method to yield the most pertinent information for each succeeding model. Each training step modifies the data distribution to minimize the error generated by the earlier models. A popular statistical learning technique called integrated learning builds a strong learner which predict the target accurately by combining several weak learners in an efficient manner. This strong learner can lower the prediction model's variance and deviation [27] [28]. Gradient Boosting algorithm is given below:

- The GBR algorithm is used to learn patterns between the input and output variables X and Y respectively, comprising N samples. The goal is to learn the function f(x) that corresponds to the mapping of input features X to target variables Y. This algorithm employs boosted trees, representing the sum of trees.

  The loss function is the difference between the actual and predicted variables, and is expressed as

$$L(f) = \sum_{i=1}^{N} L(y_i, f(x_i)) \tag{5}$$

- The minimum loss function $L(f)$ with respect to f is given as

$$\hat{f}_0(x) = \underset{f}{argmin}\ L(f) = \underset{f}{argmin} \sum_{i=1}^{N} L(y_i, f(x_i)) \tag{6}$$

  If our gradient boosting algorithm comprises M stages, then to enhance f$_m$, the algorithm can introduce a new function h$_m$, where $1 \leq m \leq M$.

$$\hat{y}_i = F_{m+1}(x_i) = F_m(x_i) + h_m(x_i) \tag{7}$$

2.3 Support Vector Regression:

Support Vector Regression (SVR) is a supervised learning method that is applicable to classification and regression tasks. It was first introduced as a linear classification technique, then extended to handle non-linear relationship and regression tasks [29]. In the SVR model, the input features X_i are transformed into a higher-dimensional feature space using a nonlinear transformation $\phi(X_i)$. A linear model is then fitted in this space, expressed as

$$E = \sum_{i=1}^{l} w_i\, \phi(X_i) + b \tag{8}$$

where w_i and b represent the learned weights and bias term respectively. The coefficients are optimized from the training dataset $\{(X_i, E_i)\}_{i=1}^{l}$. Various types of functions including gaussian, polynomial and sigmoid are employed by the SVR algorithm [30]. The fundamental concept behind the SVM method involves converting the input features into a space with higher dimensions, enabling the linear separation of the two classes through a high-dimensional surface referred to as a hyperplane. SVR mainly finds the optimal hyperplane in the transformed feature space that best fits the data [31].

2.4 CatBoost Regressor

CatBoost Regressor (CBR) is an advanced ensemble learning model based on gradient boosting, designed to improve prediction accuracy and computational efficiency. It differs from XGBoost and LightGBM, due to its construction of balanced trees that exhibit symmetry in their structure. This distinctive approach involves selecting and applying the same feature-split pair, leading to the lowest loss, across all nodes within a given level during each step [32]. The balanced tree architecture offers numerous advantages, including support for efficient CPU implementation, reduced prediction time, streamlined model application and controlled risk of overfitting. CBR is effective for a small dataset, although in such cases we should be careful about overfitting. Overfitting can be avoided by tuning the model parameters. This method can improve model correctness and speed up convergence, particularly for large feature data sets.

The prediction of the model given by Dorogush et al. [33] can be expressed as follows

$$Z = H(x_i) = \sum_{j=1}^{J} c_j 1_{\{x \epsilon R_j\}} \tag{9}$$

Here, $H(x_i)$ represents the output of the model for the input variable x_i and R_j denotes the disjoint region defined by the tree structure.

3 Results and discussion

In this study, we calculated feature importance, predicted the value of formation energy by using different machine learning models: RF, GBR, SVR and CBR. We calculated the predictive performance of these regression models using two standard metrics: coefficient of determination (R^2) and the root mean squared error (RMSE). We compared the actual and predicted value by plotting graph. Hyperparameter tuning was used to reduce overfitting and to improve models performance.

To understand the data better, we created a correlation heatmap and boxplots to examine the relationships among features. Since our data contain both experimental and theoretical material properties, no outlier removal was performed as they may represent meaningful properties rather than noise. To ensure that all features contributed equally during training, we applied z-score normalization to standardize the data. The dataset was then split into features and the target (formation energy), followed by a 80 : 20 division into training and test sets. We have used the RF model to identify important feature. From sci-kit learn, RF model assigned score to each input features. Top 10 features were selected on the basis of RF importance score. We analyzed feature importance to find which variables most influenced the formation energy predictions. These steps helped us to prepare the data for higher computational efficiency, explore its characteristics, and interpret the model's key drivers. The Random Forest model highlighted the most influential features for predicting formation energy, as shown in Figure 3.

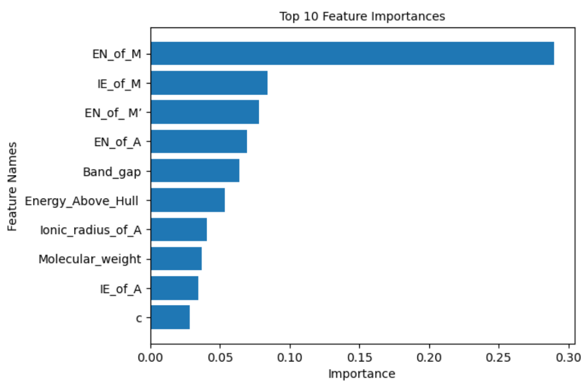

Figure 3: Top 10 features determined by the RF value importance scores analysis on the formation energy prediction.

Here, feature names are in descending order along y-axis and the importance weights along x-axis. Physically significant properties such as electronegativity, ionization energy, band gap were ranked highest, indicating strong links to material stability. Electronegativity, ionization energy and ionic radius affect bond strength and stability of molecules. The energy above hall is the key component to determine the stability of the compound, which is also identified as important by machine learning. This result supports the reliability of the model and helps narrow down key descriptors for future analysis.

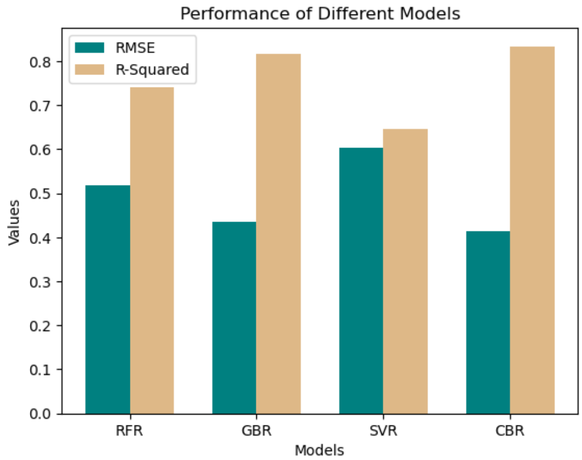

Figure 4: RMSE and R^2 value for different models.

The formation energy was predicted by using all the top features as input parameters. The most commonly tuned hyperparameter were used to evaluate the performance of different machine learning algorithms using metrics: RMSE and R^2. Reported values are based on the test dataset which evaluates the predictive performance of the model on unseen data. Table 1 shows the RMSE, R^2 values and the best parameters for hyperparameter tuning of RFR, GBR, SVR and CBR. From the above Table 1, CBR Model has RMSE : 0.4138 which is the minimum value of the root mean squared error among four models: RF, GBR, SVR and CBR. The Support Vector Regression model has the highest value of RMSE i.e. 0.6037. CatBoost Regressor model has maximum R^2 value: 0.8335 whereas the Support Vector Regression model has minimum value of R^2, i.e. 0.6456.

The CBR model achieved the highest performance among the models tested due to its ability to manage complex feature interactions and perform well with limited data. Its use of symmetric trees and ordered boosting effectively minimizes overfitting and enhances generalization. The SVR model exhibited relatively lower performance due to its smooth and symmetric feature-target relationships.

However, formation energy often involves abrupt, nonlinear variations, particularly due to structural transformation among oxides, making it challenging for SVR to model accurately. Both Gradient

Boosting Regressor and Random Forest Regressor performed reasonably well, but showed some limitations. Their sensitivity to noise and risk of overfitting, especially in the absence of robust regularization may have affected their performance on the moderately sized dataset used in this study. Monareng et al. predicted the formation energy of lithium ion batteries using CatBoost regressor model with R^2=0.95 [34]. This demonstrates the effectiveness of CBR model and suggests its applicability beyond lithium-ion battery materials. In Figure 4, among the four models, the largest coefficient of determination and the least value of the RMSE is observed in the CBR algorithm. The Support Vector Regressor algorithm has the lowest value of R^2 and the highest value of RMSE.

Table 1: Performance comparison of four different models to predict formation energy with the RMSE, R^2, hyperparameter and the optimized hyperparameter for each model

Model	RMSE	R^2	Hyperparameter	Best parameters
RFR	0.5170	0.7401	'max_depth': [3,5,7], 'n_estimators': [10,50,100]	'max_depth': 7, 'n_estimators': 100
GBR	0.4353	0.8157	'max_depth': [3,5,7], 'n_estimators': [10,50,100]	'max_depth': 3, 'n_estimators': 100
SVR	0.6037	0.6456	'C': [1,10,100], 'kernel': ['linear', 'rbf']	'C': 10, 'kernel': 'rbf'
CBR	0.4138	0.8335	'depth': [3,5,7], 'iterations': [100,200,300]	'depth': 3, 'iterations': 300

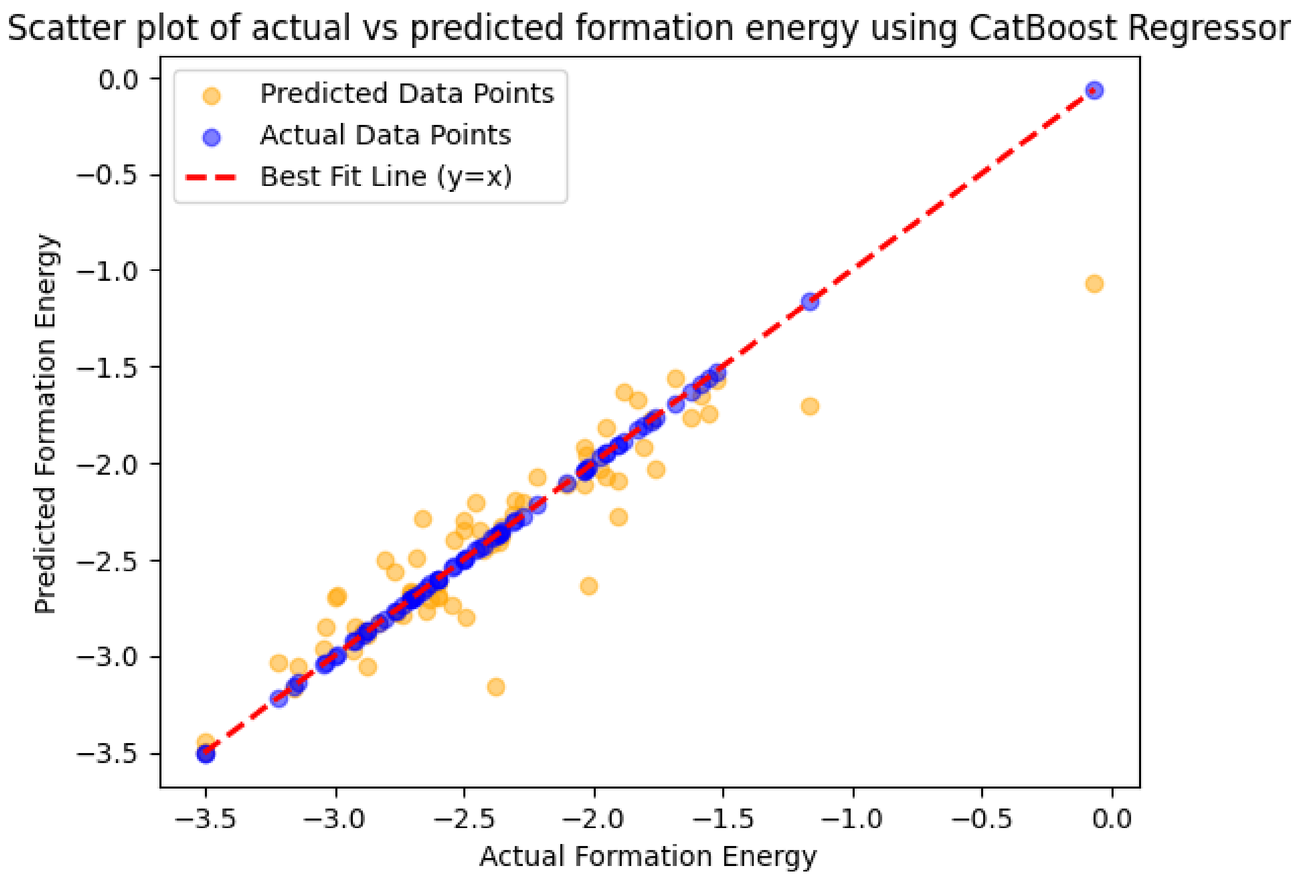

Figure 5: Scatter plot using CBR model.

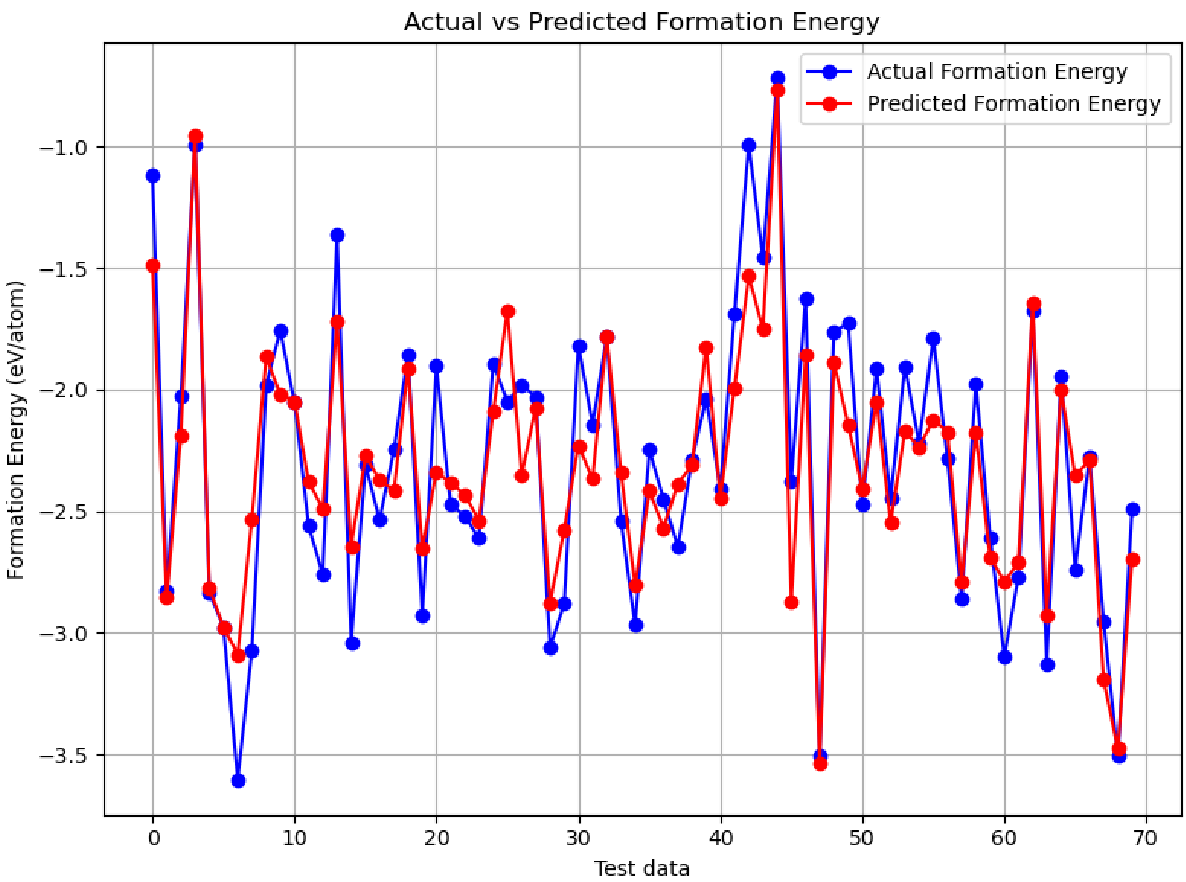

Figure 6: Visual representation of correspondence between test and predicted formation energy.

Table 2 shows the actual data and closely predicted values of the formation energy of $A_xM_yM'_zO_6$ oxides using CBR model. Figure 5, is the scatter plot of the actual and predicted data of the formation energy. We can see that there is a high correlation between the predicted and actual formation energy. The predicted and actual values of the formation energy are very close. Figure 6 further represents the comparison of the test data and predicted formation energy in a point-to-point manner to highlight the consistency of the CBR model. The close overlap between the two curves shows that the model is able to follow the actual trend of formation energy across different oxide compositions with minimal deviation.

Table 2: Actual and CBR predicted values of formation energy

Compound	Actual FE	Predicted FE
$Li_2Cr_2CoO_6$	2.052	2.047
$Na_3Fe(BO_3)_2$	2.310	2.271
$Cu_2H_3NO_6$	0.991	0.955
$LiSi_2BO_6$	3.059	2.876
$MgCr(SiO_3)_2$	2.976	2.976
Sr_2TbWO_6	3.126	2.926

The CatBoost Regressor model gives good results, but some prediction errors are still present. This may be due to the limited range of input features. Additionally, the dataset itself may include noise arising from experimental inaccuracies or inconsistencies. The presence of compounds with varying crystal symmetries could also pose challenges, as such diversity may not be fully captured by the model. Despite these challenges, predicting formation energy correctly is important to find stable oxide materials that can be used in batteries, solar cells, and catalysts.

4 Conclusion

This work highlights the potential of machine learning to predict the formation energy of oxide compounds using theoretically and experimentally collected data. The random forest regressor model identifies key descriptors such as electronegativity, ionization energy, band gap, and the energy above hull as significant factors. Based on these features, we applied four regression models to a dataset collected from the materials project. Machine learning excels at examining large datasets, uncovering trends and underlying patterns within complex information. In this study, 80% of the data was used for training, while the remaining 20% was used for the testing. Model performance evaluated using RMSE and R^2 metrics. The performance of the model was observed through the RMSE and R^2 values. The CatBoost Regressor model predicted the formation energy with higher accuracy than the other models, achieving an R^2 value of 0.83. The best hyperparameter for the prediction of formation energy is 'depth': 3, 'iterations': 300. This work presents an efficient and accelerated approach for discovering novel materials in the field of materials science. The approach used in this work can create a precise model for predicting the energy required for oxide production. The findings of this study give valuable insights for advancing the development of high-performance oxides.

Acknowledgements

We are grateful to NVIDIA for providing the Deep Learning software, which has been crucial support for our machine learning work. M.P.G. acknowledges the support of the University Grants Commission, Nepal, through the Collaborative Research Grants (Award No. CRG-78/79 S&T-03). We also acknowledge the Material Project Database for data access.

References

[1] Peng J, Jury EC, Dönnes P, Ciurtin C. Machine learning techniques for personalised medicine approaches in immune-mediated chronic inflammatory diseases: applications and challenges. Fron Pharm. 2021;12:720694.

[2] Chaudhary L, Chaudhary K, Shahi A, Jaiswal KN, Kalauni DP, Kim SH, et al. Machine learning based prediction of specific heat capacity for half-Heusler compounds. AIP Advances. 2025;15:015306.

[3] Doan T, Kalita J. Selecting machine learning algorithms using regression models. In: 2015 IEEE International Conference on Data Mining Workshop; 2015. p. 1498-505.

[4] Rashidi HH, Tran NK, Betts EV, Howell LP, Green R. Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. Acad Pathol. 2019;6:2374289519873088.

[5] Schleder GR, Padilha ACM, Acosta CM, Costa M, Fazzio A. From DFT to machine learning: recent approaches to materials science–a review. Journal of Physics: Materials. 2019 may;2(3):032001.

[6] Mueller T, Kusne AG, Ramprasad R. Machine learning in materials science: Recent progress and emerging applications. Reviews in Computational Chemistry. 2016;29:186-273.

[7] Wei J, Chu X, Sun XY, Xu K, Deng HX, Chen J, et al. Machine learning in materials science. InfoMat. 2019;1(3):338-58.

[8] Ulissi ZW, Medford AJ, Bligaard T, Nørskov JK. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. Nature Communications. 2017;8(1):14621.

[9] Mao Y, Yang H, Sheng Y, Wang J, Ouyang R, Ye C, et al. Prediction and classification of formation energies of binary compounds by machine learning: an approach without crystal structure information. ACS omega. 2021;6(22):14533-41.

[10] Zhang Z, Li M, Flores K, Mishra R. Machine learning formation enthalpies of intermetallics. J Appl Phys. 2020;128(10).

[11] Peng J, Gunda NSH, Bridges CA, Lee S, Haynes JA, Shin D. A machine learning approach to predict thermal expansion of complex oxides. Computational Materials Science. 2022;210:111034.

[12] Li W, Jacobs R, Morgan D. Predicting the thermodynamic stability of perovskite oxides using machine learning models. Computational Materials Science. 2018;150:454-63.

[13] Takahashi A, Kumagai Y, Miyamoto J, Mochizuki Y, Oba F. Machine learning models for predicting the dielectric constants of oxides based on high-throughput first-principles calculations. Physical Review Materials. 2020;4(10):103801.

[14] Mattur MN, Nagappan N, Rath S, Thomas T, et al. Prediction of nature of band gap of perovskite oxides (ABO3) using a machine learning approach. Journal of Materiomics. 2022;8(5):937-48.

[15] Tian-Xing Y, D P. Prediction of formation energy for oxides in ODS steels by machine learning. Materials Design. 2024;248:113503.

[16] Yang TX, Dou P. Prediction of formation energy for oxides in ODS steels by machine learning. Materials and Design. 2024;248:113503.

[17] Wan Z, Wang QD, Liu D, Liang J. Data-driven machine learning model for the prediction of

oxygen vacancy formation energy of metal oxide materials. Physical Chemistry Chemical Physics. 2021;23(29):15675-84.

[18] Yao W, Jia W, Shen R, Wang J, Zhang L, Wang X. Machine learning prediction of bandgap and formation energy in two-dimensional metal oxides. Physica B: Condensed Matter. 2025:417821.

[19] Pilania G, Wang C, Jiang X, Rajasekaran S, Ramprasad R. Accelerating materials property predictions using machine learning. Sci Rep. 2013;3(1).

[20] Ward L, Dunn A, Faghaninia A, Zimmermann NE, Bajaj S, Wang Q, et al. Matminer: An open source toolkit for materials data mining. Comput Mater Sci. 2018;152:60-9.

[21] Lotfi S. Targeting productive composition space through machine-learning-directed inorganic synthesis. Matter. 2020;3:261-72.

[22] https://materialsproject.org/.

[23] Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing. 2020;415:295-316.

[24] Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci. 2021;7:e623.

[25] Breiman L. Random forests. Machine Learning. 2001;45:5-32.

[26] Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas M. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol Rev. 2015;71:804-18.

[27] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001:1189-232.

[28] Wang MX, Huang D, Wang G, Li DQ. SS-XGBoost: a machine learning framework for predicting newmark sliding displacements of slopes. J Geotech Geoenviron Eng. 2020;146(9):04020074.

[29] Vapnik V, Izmailov R. Reinforced SVM method and memorization mechanisms. Pattern Recognit. 2021;119:108018.

[30] Jabeur SB, Gharib C, Mefteh-Wali S, Arfi WB. CatBoost model and artificial intelligence techniques for corporate failure prediction. Technol Forecast Soc Change. 2021;166:120658.

[31] Sánchez A VD. Advanced support vector machines and kernel methods. Neurocomputing. 2003;55(1-2):5-20.

[32] Ibrahim AA, Ridwan RL, Muhammed MM, Abdulaziz RO, Saheed GA. Comparison of the CatBoost classifier with other machine learning methods. Int J Adv Comput Sci Appl. 2020;11(11).

[33] Dorogush AV, Ershov V, Gulin A. CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:181011363. 2018.

[34] Monareng KM, Maphanga RR, Ntoahae SP. Machine Learning Model for Predicting Formation Energies for Lithium-ion Battery Materials. 2021.