INTERNATIONAL JOURNAL OF TECHNOLOGY AND ACADEMIC RESEARCH

ISSN: 2224-6878 | Pakistan Volume 6, No. 11.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

Machine-learning screening of Janus MXenes and DFT validation of their thermodynamic stability

J. Ochilov[1], A. Azamjonov[1], Sh. Khasanov[2], R. Dilmurat[4], U. Khalilov[1,3,*]

[1] Arifov Institute of Ion-Plasma and Laser Technologies, Academy of Sciences of the Republic of Uzbekistan, Tashkent, Uzbekistan

[2] Namangan State University, Namangan, Uzbekistan

[3] MOSAIC Research Group, University of Antwerp, Antwerp, Belgium

[4] King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

* Corresponding author: umedjon.khalilov@outlook.com

doi:https://doi.org/10.5281/zenodo.24513214

Abstract. The compositional space of Janus MXenes is too large for exhaustive first-principles evaluation. This work applies a two-stage approach: a machine-learning (ML) model screens the candidates, and density functional theory (DFT) validates them. The models were trained on 114,168 inorganic compounds from the Materials Project. Each compound was described by 45 Matminer descriptors. Among eleven regression algorithms, Random Forest Regression (RFR) performed best: $R^2 = 0.893$, mean absolute error 0.175 eV/atom. A set of 220 Janus MXenes of M_2XTT' composition was generated from the Zr_2COF prototype. Energetic criteria retained 48 of them, and practical criteria selected 10. Quantum ESPRESSO calculations gave negative formation energies for all ten candidates (−0.229 to −1.673 eV/atom). On a convex hull built only from Quantum ESPRESSO energies, six MXenes are stable: Ti_2CClBr, Zr_2CClBr, V_2CClBr, Nb_2CClBr, Ta_2CClBr and V_2NOBr. The ML model reproduces the ordering of the compositions by formation energy (Spearman $\rho = 0.94$). Its error is larger for nitride–oxide compositions than for carbide–halides.

Keywords: Janus MXene, machine learning, formation energy, convex hull, density functional theory, Random Forest Regression, screening.

1. Introduction

MXenes are two-dimensional transition-metal carbides and nitrides. They were discovered in 2011 [1]. MXenes are used in energy storage, electromagnetic shielding, catalysis and electronics [2]. Chemical groups (terminations) cover the MXene surface. The electronic and catalytic properties of the layer depend on these groups and on the transition metal. Besides the conventional O, OH and F groups, Cl, Br, I, S and Se terminations can now also be obtained [3].

In a Janus MXene, the two surfaces of the layer carry different groups. This asymmetry creates an intrinsic dipole in the layer. A similar structure was obtained experimentally in the

ISSN: 2224-6878 | Pakistan Volume 6, No. 11.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

MoSSe monolayer [4]. Calculations show that Janus Zr_2COT (T = F, Cl, Br, I) MXenes have higher hydrogen evolution reaction (HER) activity than symmetric structures [5].

The number of possible Janus compositions reaches several hundred. A DFT calculation of each composition requires large computational resources. Therefore, candidates are first screened by an inexpensive ML model. Only the selected candidates are then checked by DFT. ML models are trained on open databases such as the Materials Project [6]. Composition-based descriptors predict the formation energy well [7, 8].

Two energetic quantities are used in screening. The first is the formation energy E_f. It shows the energetic preference of a compound relative to the elements. A negative E_f, however, does not guarantee stability. A compound may decompose into competing phases of lower energy. The second is the energy above the convex hull E_{hull}. It measures stability against exactly this decomposition. For most experimentally obtained metastable compounds, E_{hull} does not exceed a few tens of meV/atom [9].

This work carries out a complete workflow for Janus MXenes, from ML screening to DFT validation. It also assesses the accuracy of the ML predictions. The paper reports four results: (1) the best regression model; (2) the screening of 220 compositions down to 10 candidates; (3) the DFT stability of the candidates; (4) the comparison of the ML predictions with DFT.

2. Methods

2.1. Training data and descriptors

Training data were taken from the Materials Project database [6]. The initial set contained 154,879 inorganic compounds. The E_f and E_{hull} values are known for each compound. Compounds with noble gases (He, Ne, Ar) and with actinides lacking complete elemental data (Np, Pu) were removed. Duplicate entries were also removed. As a result, 122,585 compounds remained.

Compositions were converted into descriptors with the Matminer library [7, 8]. The descriptors describe atomic size, chemical bonding, electronic structure, valence configuration and composition. After entries with incomplete descriptors were removed, 114,168 compounds and 45 descriptors remained. E_f and E_{hull} were treated as two independent regression targets.

2.2. Candidate set

The Zr_2COF Janus MXene from the aNANt database was taken as the prototype [10]. The element combinations followed Ref. [5]. Candidates are described by the formula M_2XTT'. Here M is a transition metal (Sc, Y, Ti, Zr, Hf, V, Nb, Ta, Cr, Mo, W), X is C or N, and T and T′ are two different terminations from the series O, F, Cl, Br, I. Five terminations give ten distinct pairs. The set thus contains $11 \times 2 \times 10 = 220$ compositions. Substitution preserved the P3m1 symmetry of the prototype (space group No. 156).

2.3. Regression models

Eleven regression algorithms from the scikit-learn library were tested [11]: the mean-value model (DUMR, baseline), Lasso (LAS), AdaBoost (ABR), linear support vector regression (LSVR), Ridge (RR), gradient boosting (GBR), support vector regression (SVR), k-nearest neighbors (KNR), multilayer perceptron (MLP), Extra Trees (ETR) and Random Forest (RFR) [12]. Descriptors were standardized before training.

ISSN: 2224-6878 | Pakistan Volume 6, No. 11.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

Model quality was assessed by the coefficient of determination R^2, the mean absolute error (MAE) and the root-mean-square error (RMSE). The assessment was performed on the training set and by five-fold cross-validation. The agreement in ordering between ML and DFT values was measured by the Spearman rank correlation coefficient ρ.

2.4. DFT calculations

Calculations were performed with Quantum ESPRESSO [13, 14]. The PBE exchange–correlation functional [15] and ultrasoft pseudopotentials from PSlibrary 1.0.0 [16] were used. The kinetic-energy cutoff is 50 Ry for the wavefunctions and 450 Ry for the charge density. The Gaussian smearing width is 0.01 Ry.

Each MXene was modeled as a 3 × 3 × 1 supercell (45 atoms). The vacuum layer corresponds to c = 31.154 Å. The Brillouin zone was sampled with Γ-centered Monkhorst–Pack grids [17]: 3 × 3 × 1 for relaxation and 15 × 15 × 1 for the final static calculation. Structures were relaxed with the BFGS algorithm in the two-dimensional variable-cell mode. The convergence thresholds were 10^{-6} Ry for the self-consistent energy, 10^{-5} Ry for the energy between ionic steps, and 10^{-4} Ry/bohr for the forces.

Elemental energies were computed with the same settings. Solids (C, Ti, Zr, V, Nb, Ta, Cr, Mo) were fully relaxed. Cl_2, Br_2, N_2 and O_2 were computed as isolated molecules. Cr was treated in its antiferromagnetic state and O_2 in its triplet state, both with spin polarization [18].

2.5. Formation energy and convex hull

The formation energy per atom was calculated as:

$E_f = [E_{tot} - \Sigma_i n_i\mu_i] / \Sigma_i n_i$

Here E_{tot} is the total energy of the compound, n_i is the number of atoms of element i, and μ_i is the chemical potential of that element.

Stability was assessed against competing phases. For each of the ten quaternary systems, the competing phases were taken from the Materials Project database [6]. Their energies were recalculated in Quantum ESPRESSO with the same settings as the MXenes. All energies are therefore mutually consistent. Of the 57 phases, 53 converged and entered the hull. Four phases ($MoCl_6$, Ta_2Br_5, $Ti_3N_2O_3$, ZrC) did not converge and were excluded.

The convex hull was built with the pymatgen library [19]. For an MXene on the hull, the stability margin was also determined. It shows how far the MXene lies below the hull built from the competing phases alone. The classification is as follows: $E_{hull} \leq 0.005$ eV/atom — stable; up to 0.050 eV/atom — near-stable; up to 0.100 eV/atom — possibly synthesizable; higher — unstable [9].

3. Results and discussion

The workflow consists of five stages: training data, candidate set, ML screening, candidate selection and DFT validation (Fig. 1).

ISSN: 2224-6878 | Pakistan Volume 6, No. 11.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

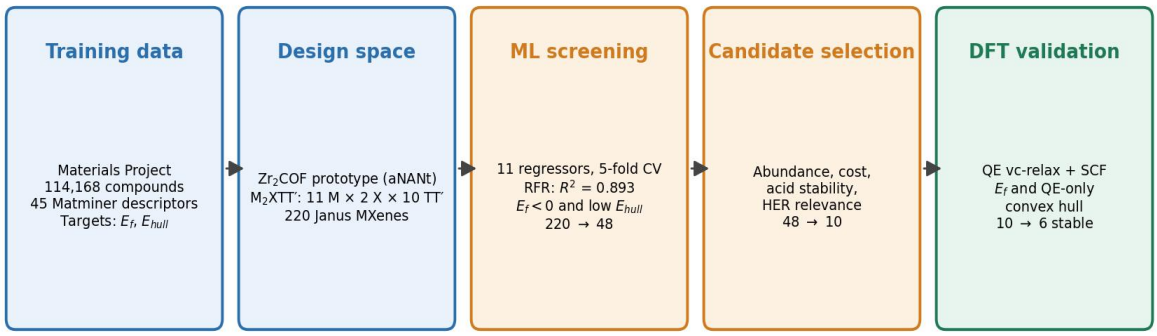

Fig. 1. Screening and validation workflow.

3.1. The best model: Random Forest Regression

RFR predicted the formation energy most accurately. In cross-validation it gave $R^2 = 0.893$, MAE = 0.175 eV/atom and RMSE = 0.397 eV/atom (Table 1, Fig. 2).

The eleven algorithms fall into three groups. DUMR, LAS and ABR have no predictive power ($R^2 \approx 0$). The linear LSVR and RR models explain about three quarters of the variance ($R^2 = 0.742$ and 0.756). The nonlinear GBR, SVR, KNR and MLP models are more accurate: $R^2 = 0.847–0.874$. Tree ensembles gave the highest scores: 0.883 for ETR and 0.893 for RFR.

Model selection also considered the gap between training and validation scores. ETR reproduces the training set almost perfectly ($R^2 = 0.983$). Its validation score, however, is lower than that of RFR. This is a sign of overfitting. For RFR the gap is smaller: 0.081 (0.100 for ETR). RFR was therefore chosen as the screening model. The same procedure also selected RFR for E_{hull}.

Table 1. Performance of the eleven regression models for the formation energy. MAE and RMSE are in eV/atom. Validation is five-fold cross-validation.

Model	Training R^2	Validation R^2	Validation MAE	Validation RMSE
DUMR	0.000	−0.000	1.011	1.211
LAS	0.000	−0.000	1.011	1.211
ABR	−0.052	−0.002	1.094	1.210
LSVR	0.745	0.742	0.395	0.615
RR	0.756	0.756	0.410	0.598
GBR	0.849	0.847	0.286	0.474
SVR	0.860	0.856	0.201	0.460
KNR	0.917	0.870	0.218	0.436
MLP	0.885	0.874	0.247	0.431
ETR	0.983	0.883	0.164	0.414
RFR	**0.974**	**0.893**	**0.175**	**0.397**

ISSN: 2224-6878 | Pakistan Volume 6, No. 11.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

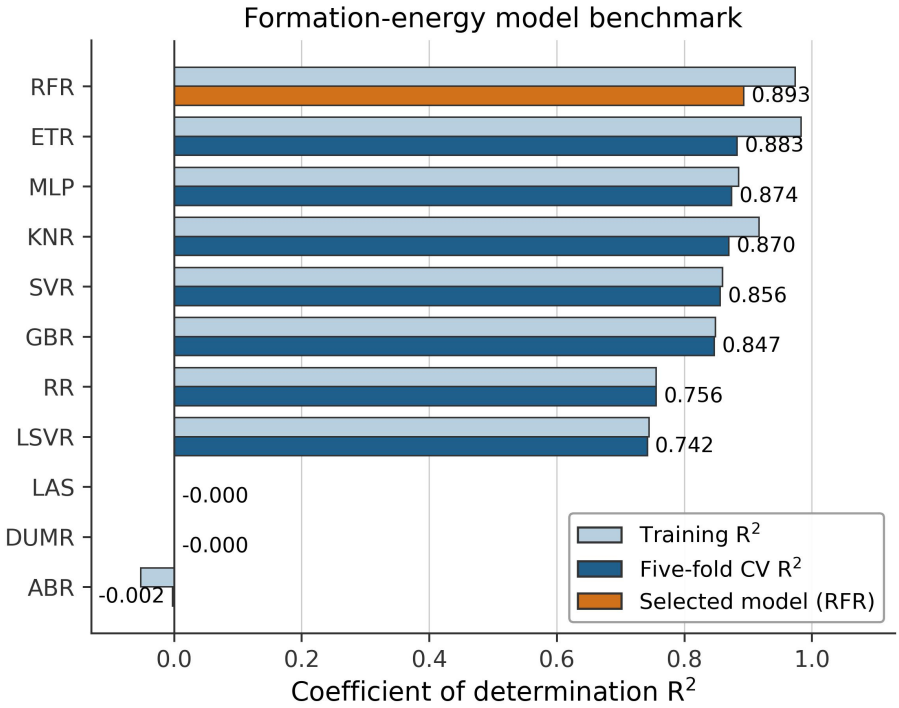

Fig. 2. R^2 values of the eleven regression models. Light blue bars: training set. Dark blue bars: five-fold cross-validation. Orange bar: the selected RFR model.

3.2. From 220 compositions to 10 candidates

Energetic criteria reduced the 220 compositions to 48, and practical criteria reduced them to 10.

The RFR model predicted a negative E_f for 208 of the 220 compositions (94.5%). Thus, E_f separates the candidates weakly in this set. The E_{hull} model gives a sharper cut. Applying both criteria together left 48 candidates. They belong to ten transition metals: Sc, Y, Ti, Zr, Hf, V, Nb, Ta, Cr, Mo. Compositions with mixed Cl/Br terminations dominate among them.

Ten of the 48 candidates were selected by three practical criteria: elemental abundance and cost, stability in acidic media, and relevance to HER. Seventeen Sc- and Y-based structures were removed because of limited reserves. Hf-based candidates were removed because of their cost; the chemically similar Zr analogues were kept. Iodine-terminated compositions were removed because of their electrochemical instability under HER conditions.

The ten selected compositions fall into two families. Carbide–halides: Ti_2CClBr, Zr_2CClBr, V_2CClBr, Nb_2CClBr, Ta_2CClBr, Cr_2CClBr and Mo_2CClBr. Nitride–oxides: Ti_2NOCl, V_2NOBr and Ta_2NOBr.

3.3. DFT validation: six stable MXenes

All ten candidates are energetically favorable relative to the elements. However, only six of them lie on the convex hull (Table 2, Fig. 3).

During relaxation, all candidates kept the five-plane T–M–X–M–T′ Janus structure. The lattice constant lies in the range 3.077–3.475 Å. DFT gives E_f values from −0.229 eV/atom (Cr_2CClBr) to −1.673 eV/atom (Ti_2NOCl). The ML model had predicted a negative E_f for all ten candidates. DFT confirmed this sign in every case.

ISSN: 2224-6878 | Pakistan Volume 6, No. 11.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

Six MXenes lie on the convex hull: Ti_2CClBr, Zr_2CClBr, V_2CClBr, Nb_2CClBr, Ta_2CClBr and V_2NOBr. Their reliability differs. Zr_2CClBr has a large stability margin: 0.276 eV/atom. Because the ZrC phase did not converge, this value is an upper bound. Ti_2CClBr (0.017 eV/atom), V_2CClBr (0.024 eV/atom) and Nb_2CClBr (0.048 eV/atom) lie on the hull with a small margin. The margins of Ta_2CClBr (0.577 eV/atom) and V_2NOBr (0.713 eV/atom) appear large. However, only one and two competing phases, respectively, converged in their systems. These two results are therefore preliminary.

Four candidates lie above the hull and decompose into stable binary compounds:

– Cr_2CClBr (E_{hull} = 0.211 eV/atom) → $Cr_3C_2 + CrCl_3 + CrBr_3 + C$;

– Mo_2CClBr (0.248 eV/atom) → $MoC + MoBr_2 + MoCl_2$;

– Ti_2NOCl (0.427 eV/atom) → $TiN + Ti_2O_3 + TiCl_3$;

– Ta_2NOBr (0.438 eV/atom) → $Ta_2N + Ta_2O_5 + TaN + TaBr_5$.

The case of Ti_2NOCl is important. Its E_f is the most negative in the set (−1.673 eV/atom). Nevertheless, it is unstable against decomposition. A negative E_f thus indicates chemical plausibility, not stability. Stability is determined by convex-hull analysis.

Table 2. Lattice constant a, ML and DFT values of E_f and E_{hull} (eV/atom), and the DFT verdict for the ten Janus MXenes. Compositions are ordered by increasing ML E_{hull}.

Composition	a (Å)	E_f (ML)	E_f (DFT)	E_{hull} (ML)	E_{hull} (DFT)	DFT verdict
Ti_2CClBr	3.279	−1.333	−1.268	0.030	0.000	On hull
Zr_2CClBr	3.475	−1.552	−1.242	0.032	0.000	On hull
Mo_2CClBr	3.383	−0.721	−0.366	0.083	0.248	Above hull
V_2CClBr	3.244	−0.737	−0.755	0.089	0.000	On hull
Cr_2CClBr	3.201	−0.638	−0.229	0.092	0.211	Above hull
Nb_2CClBr	3.417	−0.806	−0.894	0.102	0.000	On hull
Ta_2CClBr	3.403	−0.771	−0.834	0.114	0.000	On hull
V_2NOBr	3.077	−1.458	−0.918	0.180	0.000	On hull
Ti_2NOCl	3.079	−1.915	−1.673	0.193	0.427	Above hull
Ta_2NOBr	3.226	−1.311	−1.171	0.194	0.438	Above hull

ISSN: 2224-6878 | Pakistan Volume 6, No. 11.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

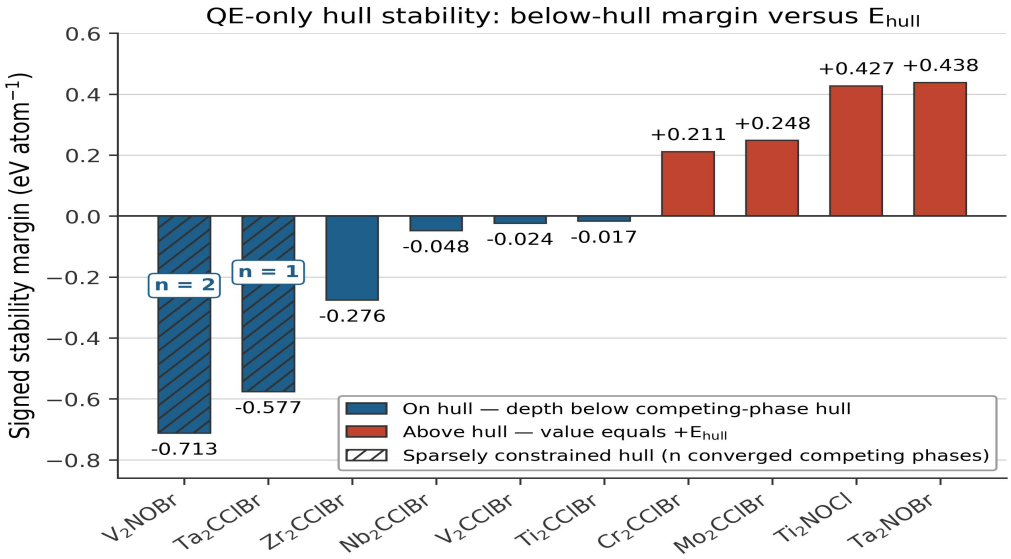

Fig. 3. Stability of the ten Janus MXenes relative to the convex hull (eV/atom). Blue bars: compositions on the hull; the bar shows the stability margin. Red bars: compositions above the hull; the bar equals E_{hull}. Hatched bars: systems with few competing phases (n is the number of converged phases).

3.4. Accuracy of the ML predictions

The ML model gives the correct ordering of the compositions by E_f but shifts the absolute values (Fig. 4).

For the ten candidates, the mean absolute deviation between ML and DFT is 0.223 eV/atom (RMSE = 0.278 eV/atom). The Spearman coefficient is $\rho = 0.94$. The shift is systematic: for seven of the ten compositions, the ML value is more negative than the DFT value. The reason is that three-dimensional bulk phases dominate the training set. The best agreement was found for V_2CClBr (deviation 0.018 eV/atom), Ta_2CClBr (0.063 eV/atom) and Ti_2CClBr (0.065 eV/atom). The largest deviations are for V_2NOBr (0.540 eV/atom) and Cr_2CClBr (0.409 eV/atom).

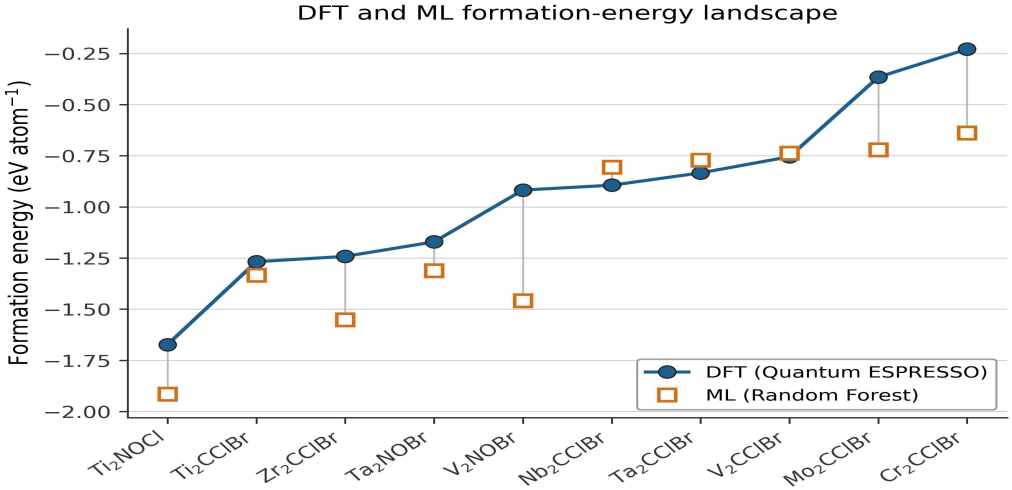

Fig. 4. DFT and ML formation energies of the ten Janus MXenes. Compositions are ordered by increasing DFT value. Blue circles: DFT (Quantum ESPRESSO). Orange squares: ML (RFR). Grey lines: the deviation between them.

ISSN: 2224-6878 | Pakistan Volume 6, No. 11.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

The E_{hull} model correctly separates only the extreme cases. The two compositions that the model ranks as most stable (Ti_2CClBr and Zr_2CClBr) lie on the DFT hull. The two compositions that it ranks as least stable (Ti_2NOCl and Ta_2NOBr) are also the least stable by DFT. The ordering of the intermediate compositions carries no information: within the carbide–halides, $\rho \approx -0.05$. The predicted range (0.030–0.194 eV/atom) is narrower than the DFT range (0.000–0.438 eV/atom). The reason is as follows. E_{hull} depends on all competing phases in the system. Composition-based descriptors do not contain this information.

The error is larger for the nitride–oxide compositions (Table 3). Such mixed two-dimensional compositions are rare in the Materials Project database. The error therefore comes from the composition of the training data, not from the algorithm. Additional training on MXene-specific data is expected to reduce this error.

Table 3. Mean absolute deviation of the ML predictions from DFT (eV/atom).

Composition family	**Number of compositions**	**E_f**	**E_{hull}**
Carbide–halides	7	0.187	0.093
Nitride–oxides	3	0.307	0.219
All compositions	10	0.223	0.131

Thus, the composition-based ML model does not replace DFT. It shortens the candidate list quickly and at low cost. Final stability is determined by full convex-hull construction.

4. Conclusions

1. Among eleven algorithms, RFR gave the best prediction of the formation energy: $R^2 = 0.893$, MAE = 0.175 eV/atom.

2. Of the 220 M_2XT' Janus MXenes generated from the Zr_2COF prototype, energetic criteria selected 48 and practical criteria selected 10.

3. DFT confirmed a negative E_f for all ten candidates. Six of them lie on the convex hull: Ti_2CClBr, Zr_2CClBr, V_2CClBr, Nb_2CClBr, Ta_2CClBr and V_2NOBr. The most reliable candidate is Zr_2CClBr (stability margin 0.276 eV/atom). Four candidates decompose into binary compounds.

4. The ML model gives the correct E_f ordering ($\rho = 0.94$). For E_{hull}, the model separates only the extreme cases. The error is larger for nitride–oxide compositions than for carbide–halides.

The six stable compositions form the basis for further studies: phonon analysis, ab initio molecular dynamics and assessment of hydrogen-storage properties.

References

1. M. Naguib, M. Kurtoglu, V. Presser, J. Lu, J. Niu, M. Heon, L. Hultman, Y. Gogotsi, M.W. Barsoum, Two-dimensional nanocrystals produced by exfoliation of Ti_3AlC_2, Adv. Mater. 23 (2011) 4248–4253.
2. A. VahidMohammadi, J. Rosen, Y. Gogotsi, The world of two-dimensional carbides and nitrides (MXenes), Science 372 (2021) eabf1581.

ISSN: 2224-6878 | Pakistan Volume 6, No. 11.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

3. H. Ding, Y. Li, M. Li, K. Chen, K. Liang, G. Chen, J. Lu, J. Palisaitis, P.O.Å. Persson, P. Eklund, L. Hultman, S. Du, Z. Chai, Y. Gogotsi, Q. Huang, Chemical scissor-mediated structural editing of layered transition metal carbides, Science 379 (2023) 1130–1135.
4. A.-Y. Lu, H. Zhu, J. Xiao, C.-P. Chuu, Y. Han, M.-H. Chiu, C.-C. Cheng, C.-W. Yang, K.-H. Wei, Y. Yang, Y. Wang, D. Sokaras, D. Nordlund, P. Yang, D.A. Muller, M.-Y. Chou, X. Zhang, L.-J. Li, Janus monolayers of transition metal dichalcogenides, Nat. Nanotechnol. 12 (2017) 744–749.
5. J. Hu, M. Fan, R. Zhang, X. Ji, L. Miao, J. Jiang, Breaking symmetry: enhanced hydrogen evolution reaction performance of Janus Zr_2COT (T = F, Cl, Br, I) MXenes by density functional theory, ACS Appl. Nano Mater. 7 (2024) 9294–9304.
6. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, APL Mater. 1 (2013) 011002.
7. L. Ward, A. Dunn, A. Faghaninia, N.E.R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K. Persson, G.J. Snyder, I. Foster, A. Jain, Matminer: An open source toolkit for materials data mining, Comput. Mater. Sci. 152 (2018) 60–69.
8. L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, npj Comput. Mater. 2 (2016) 16028.
9. W. Sun, S.T. Dacek, S.P. Ong, G. Hautier, A. Jain, W.D. Richards, A.C. Gamst, K.A. Persson, G. Ceder, The thermodynamic scale of inorganic crystalline metastability, Sci. Adv. 2 (2016) e1600225.
10. aNANt: A functional materials database, Materials Research Centre, Indian Institute of Science, http://anant.mrc.iisc.ac.in/ (accessed July 2026).
11. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
12. L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
13. P. Giannozzi, S. Baroni, N. Bonini, et al., QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, J. Phys.: Condens. Matter 21 (2009) 395502.
14. P. Giannozzi, O. Andreussi, T. Brumme, et al., Advanced capabilities for materials modelling with Quantum ESPRESSO, J. Phys.: Condens. Matter 29 (2017) 465901.
15. J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (1996) 3865–3868.
16. A. Dal Corso, Pseudopotentials periodic table: From H to Pu, Comput. Mater. Sci. 95 (2014) 337–350.
17. H.J. Monkhorst, J.D. Pack, Special points for Brillouin-zone integrations, Phys. Rev. B 13 (1976) 5188–5192.

ISSN: 2224-6878 | Pakistan Volume 6, No. 11.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

18. L. Wang, T. Maxisch, G. Ceder, Oxidation energies of transition metal oxides within the GGA+U framework, Phys. Rev. B 73 (2006) 195107.
19. S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, Comput. Mater. Sci. 68 (2013) 314–319.