IOP Publishing J. Phys. Energy 5 (2023) 034005 https://doi.org/10.1088/2515-7655/acb2f8

Journal of Physics: Energy

OPEN ACCESS

RECEIVED
12 June 2022

REVISED
15 December 2022

ACCEPTED FOR PUBLICATION
13 January 2023

PUBLISHED
9 May 2023

PAPER

Predicting the work function of 2D MXenes using machine-learning methods

Pranav Roy[1,5], **Lavie Rekhi**[1,5], **See Wee Koh**[2,3], **Hong Li**[2,3,*] **and Tej S Choksi**[1,4,*]

[1] School of Chemical and Biomedical Engineering, Nanyang Technological University, 62 Nanyang Drive, Singapore 637459, Singapore
[2] School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore
[3] CNRS-International-NTU-THALES Research Alliances/UMI 3288, Research Techno Plaza, 50 Nanyang Drive, Singapore 637553, Singapore
[4] Cambridge Centre for Advanced Research and Education in Singapore, CREATE Tower, 1 Create Way, Singapore 138602, Singapore
[5] These authors contributed equally to this work.
[*] Authors to whom any correspondence should be addressed.

E-mail: ehongli@ntu.edu.sg and tej.choksi@ntu.edu.sg

Keywords: MXenes, work function, machine learning, density functional theory, two-dimensional materials

Supplementary material for this article is available online

Abstract

MXenes, which are graphene-like two-dimensional transition metal carbides and nitrides, have tunable compositions and exhibit rich surface chemistry. This compositional flexibility has resulted in exquisitely tunable electronic, optical, and mechanical properties leading to the applications of MXenes in catalysis, electronics, and energy storage. The work function of MXenes is an important fundamental property that dictates the suitability of MXenes for these applications. We present a series of machine learning models to predict the work function of MXenes having generic compositions and containing surfaces terminated by O*, OH*, F*, and bare metal atoms. Our model uses the basic chemical properties of the elements constituting the MXene as features, and is trained on 275 data points from the Computational 2D Materials Database. Using 15 different features of the MXene as inputs, the neural network model predicts the work function of MXenes with a mean absolute error of 0.12 eV on the training data and 0.25 eV on the testing data. Our feature importance analysis indicates that properties of atoms terminating the MXene surface like their electronegativity, most strongly influence the work function. This sensitivity of the work function to the surface termination is also elucidated through experimental measurements on Ti_3C_2. We introduce reduced-order models comprising of ten-, eight-, and five-features to predict the work function. These reduced-order models exhibit easier transferability to new materials, while exhibiting a marginal increased mean average error. We demonstrate the transferability of these reduced order models to new materials, by predicting the work function of MXenes having surface terminations beyond the original training set, like Br*, Cl*, S*, N*, and NH*. Predicting electronic properties like the work function from the basic chemical properties of elements, paves the way towards rapidly identifying tailored MXenes having a targeted range of properties that are required for a specific application.

1. Introduction

Two-dimensional carbides and nitrides, called MXenes, are an emerging class of low-dimensional materials that have received considerable interest over the last decade [1, 2]. MXenes have a general chemical formula of $M_{n+1}X_nT_x$, with n ranging from 1 to 3. Here, M is an early transition metal (e.g. Ti), X is carbon or nitrogen, and T_x are the functional groups that terminate the surface. As-synthesized MXenes expose surfaces that are terminated by either bare metal atoms, O*, OH*, or F*. Recent studies have expanded the range of possible surface terminations of MXenes to include Cl*, S*, Br*, Te*, Se*, and NH* [3]. Such terminations are incorporated onto the surface through annealing treatments and substitution reactions [3].

By systematically altering the surface chemistry of MXenes [4, 5], we can readily tune electronic properties like the electronic band gap, the work function, density of states at the Fermi level, and the electrical conductivity [6]. Such exquisitely tunable electronic properties of MXenes have led to their applications in catalysis [7–10], energy storage [2, 11], photonics [12], and in electromagnetic shielding [13, 14].

The work function is an electronic property of a material that represents the minimum energy required to remove an electron from the Fermi level to vacuum. Aside from being a property of fundamental interest in surface science, the work function of an MXene is a metric that has relevance in catalysis, energy storage/conversion, and electronics. MXenes with high work function are favoured in optoelectronics [15, 16] and sensing while those with low work function values find applications in electronics [17, 18]. MXenes are increasingly used as supports for noble metal catalysts [19–21] because of their highly tunable surface chemistry [22–24]. One factor that governs the stability and reactivity of such supported metal catalysts is the charge transfer between the metal and support [25–27]. The work function serves as an indicator for both the extent of charge transfer and the dipole moment at the metal/support interface [27, 28]. The work function is also indirectly related to the rates of catalytic cycles through its dependence on the Fermi level of the system. This is because, in an elementary step of a catalytic cycle, the rate of electron transfer between a reaction intermediate and the catalyst depends on the Fermi level of the solid.

The interplay between surface terminal groups of MXenes and their local dipoles impacts the electronic properties of MXenes [29]. The nature of the metal (M) and the number of layers determines the terminations of a given MXene [30]. Previous experimental and computational studies have shown how changing the dominant surface termination (from among O^*, OH^*, and F^*) alters the work function [31–33]. For instance, OH^* terminated MXenes such as M_2C and $M_{10}C_9$ (where M = Sc, V, Ti, Ta, Nb, Hf, Zr) as well as M'_2N and $M'_{10}N_9$ (where M′ = Ti, Hf, Zr) possess lower work function as compared to the as-synthesized O^*/F^* terminated MXenes [31]. The decrease in the work function occurs because of the influence of the intrinsic dipole moment of the hydroxyl group. The high work function for surfaces terminated by O^* and F^* is due to charge transfer at these sites [32]. In addition to redistributions of the electronic charges transferred between the surface terminations and the sub-surface atoms, contributions of the induced dipole moments from surface relaxations and the dipole moment of adsorbates must also be considered. The dipole moment of OH^* is opposite in sign to the induced dipole moment from structural relaxation effects, resulting in OH^* terminated MXenes having lower work function values. On the other hand, trends in the work function of MXenes terminated with O^* and F^* are more sensitive to the transition metal identity [31]. Computational studies typically employ first principles methods like density functional theory (DFT) to determine electronic structure properties such as work function. It is however a computationally intensive process to determine these electronic properties from first principles for the wide range of materials within the MXenes family.

The advent of databases [34–36] and machine learning methods [37–51] have enabled high-throughput searches for targeted electronic properties of materials across a wide-range of materials. For instance, databases like the Computational 2D Materials Database (C2DB) [52, 53] and 2DMatPedia [54] store thermodynamic, structural, and electronic properties of materials like MXenes. These databases can also function as training sets for building machine learning models that estimate electronic properties [55], which are otherwise determined from DFT calculations. For instance, Hashimoto *et al* [56] used bulk properties obtained from the Materials Project database to train a Gaussian Process Regression model. This model estimated a quick measure of the work function, which was based on the depth of the Fermi level. This quick measure of the work function, though approximate, enabled the rapid identification of bulk compositions having extreme values of the work function. Rajan *et al* regressed a set of features describing MXenes to the electronic band gap. These features included boiling/melting points, atomic radii, and bond lengths [57]. Regression models like the ones discussed above are more powerful if the feature set can be constructed without additional DFT calculations. Examples of such features include the structural details of a material and physico-chemical properties of the constituting elements. For any new material outside the original training set, the above-mentioned features can be determined from a database, on-the-fly. Inputting these features into the machine learning model thus enables the direct calculation of electronic properties for the new material.

Regression-based machine learning models not only predict various properties [58] of materials relevant towards applications in catalysis, energy, and in electronic devices, but they also provide physical insights through feature importance analyses [57, 59, 60]. Lamoureux *et al* used a feature importance analysis combined with a genetic algorithm to understand how the features that describe the stability of catalytic nanoparticles evolve with the nanoparticle diameter. In the limit of large nanoparticles, they showed that the most important features identified by the machine learning model resemble those used by a physics-based linear model [61]. Rajan *et al* [57] showed that the most important features for predicting the electronic

band gap of MXenes are the volume per atom, mean boiling point of the constituting elements, and the standard deviation of the group number of the constituting elements.

While databases storing the electronic properties of MXenes have emerged, predictive models that estimate the work function of any generic MXene from pre-tabulated features are still lacking. Furthermore, our understanding regarding which physico-chemical properties lead to the exquisitely tunable work function of MXenes is still in its nascent stage. To fill these knowledge gaps, it is necessary to build a computational framework that not only predicts the work function of MXenes, but also explains how variations in the MXene composition influences the work function. Such a framework can serve as a powerful tool for designing MXenes in different catalysis and clean energy applications where the work function is an important performance metric.

Herein, we present a series of machine learning models that predict the work function of MXenes from a feature set that is primarily composed of the physico-chemical properties of their constituting elements. Our models are trained using 275+ datapoints from C2DB [52, 53]. The sensitivity of the work function to surface terminations as seen from the model is confirmed through experiments. We use three complementary methods to identify the most important physico-chemical properties that govern the work function values. These methods include: determining feature importance, understanding the occurrence probability of features in the best performing models, and using a symbolic transformer within a genetic programme. We finally elucidate how our models can be efficiently retrained as the database is augmented with new MXene compositions.

2. Methods

2.1. Computational details for the machine learning models

We consider a broad set of MXenes from the C2DB [52, 53] having chemical compositions of M_2X, M_3X_2 and M_4X_3. These MXenes are terminated with either the bare metal atoms, O^*, F^*, or OH^*. We construct a high-dimensional feature space using physico-chemical properties of the elements constituting the MXenes. These properties can be obtained from databases [62–64] and do not require additional first principles calculations. The feature set uses both physical (e.g. orbital radius of the outermost orbitals) and electronic (e.g. ionisation potential) features. Additional features related to bulk properties (e.g. lattice constant) and the thermodynamic stability (e.g. energy above convex hull, heat of formation) are also extracted from C2DB and used in the feature set. We note that these three properties are obtained from DFT calculations. Thus, our primary dataset consists of 15 features. These features are listed in tables S2 in the supporting information. Prior to implementing the dataset in our machine learning models, the features are normalized by transmuting the features such that the values lie between 0 and 1. This transmutation is done to ensure a consistent distribution of the features, regardless of their units and magnitude. Next, we construct a Pearson correlation for the primary features (figure S1 in the supporting information).

We also construct a series of reduced-order feature sets containing less than 15 features. These reduced-order feature sets will mitigate the risk of overfitting and concurrently aid in isolating the most critical features that govern the work function. Based on the Pearson correlation matrix for the 15-feature dataset, we first create a reduced-order dataset that removes those features exhibiting greater than 75% correlation with each other. Furthermore, we ensure that the reduced-order feature set includes at least one feature specific to the metal (M), carbon/nitrogen (X), termination (T) and structure of the MXenes ($M_{n+1}X_nT_x$). This process yields a ten-feature dataset. We note however, that two of the ten features (heat of formation and energy above the convex hull) are determined through structural optimization calculations performed using DFT. We thus create an eight-feature dataset that only contains those features that can be obtained without additional DFT calculations. We then leverage our understanding from the eight- and ten-feature models to build more minimalistic five-feature datasets and verify whether the model accuracy is retained.

We employ several supervised machine learning algorithms like ordinary linear regression, random forest models, and neural networks to predict the work function of MXenes. The scikit-Learn Python package [65] is used to deploy the linear regression and random forest models. The models are evaluated using the mean absolute error (MAE) and the coefficient of determination (R^2). The MAE is the average of the absolute differences between the model predicted value and the value obtained from C2DB [52, 53]. The MAEs in the training and testing set are compared to ensure that over-fitting is reduced to the maximum extent possible. In addition to applying simple linear regression models, we also apply linear regression models which are regularized using Lasso and Ridge regularizations. We further apply the random forest model with the only hyperparameter being the 'criterion', which is set to the MAE. The deep learning models are deployed using the Keras/TensorFlow package [66]. The implemented feed-forward neural network model consists of an input layer, multiple hidden layers comprising of dense, batch normalization, and dropout layers, and an

output layer. Multiple activation functions such as relay linear unit and the sigmoid function are applied in the dense layers. The sigmoid activation allows the neural network model to account for non-linearity and the model is trained to optimize the loss function, which is the MAE. The hyperparameters used in the neural network model are listed in table S4 of the supporting information.

An evolutionary algorithm called genetic programming is implemented via the gplearn module [67] to construct analytical forms of complex meta-features from individual features. Such analytical forms increase the interpretability of our model. The symbolic transformer attribute in gplearn uses a list of primary features as inputs to create a random formula called a program or a meta-feature. Each meta-feature represents a complex mathematical relationship between the primary features. This meta-feature is constructed using different mathematical operators. Further details are in tables S5 and S6 of the supporting information. The meta-features are evaluated based on the Pearson correlation coefficient with the work function (target variable) being the fitness metric. In each iteration of the genetic programme, called a generation, a population of n programs (where $n = 1000$) is created. From this population, the top x programs (where $x = 20$) in terms of fitness metric compete to become part of the next generation. The selected programs undergo multiple genetic operations such as crossover, mutation, or reproduction to improve the fitness, following which the process is repeated until the final generation. In the final generation, the best performing programs are chosen from a group of fittest programs, known as hall of fame. This hall of fame finally provides us with the meta-features that are strongly correlated to the work function. We have uploaded all relevant python scripts used to train the model on GitHub: (https://github.com/tejchoksi/workfunction-MXenes).

2.2. First principles calculations

We perform DFT calculations on MXenes having surface terminations that are different from the MXenes within the C2DB [52, 53, 68]. These MXenes include Ti_3C_2, Ti_3N_2, Nb_3C_2, Nb_3N_2, Mo_3C_2, Mo_3N_2, V_2C, and Ti_2N that are terminated with Cl^*, Br^*, S^*, N^*, and NH^*. These additional terminations are selected following recent experimental reports discussing their synthesis [3, 69]. The DFT calculations are carried out using the Quantum ESPRESSO package [70] within the Atomic Simulation Environment [71]. We employ the Perdew–Burke–Ernzerhof (PBE) exchange-correlation functional [72]. Core states of atoms are represented using Vanderbilt ultra-soft pseudopotentials [73]. The Kohn–Sham equations are solved using a plane-wave basis set, with an energy cutoff of 500 eV and a density cut-off of 5000 eV. The Kohn Sham equations are solved with convergence thresholds of 10^{-7} eV for the total energy and 0.05 eV $Å^{-1}$ for the maximum net force per atom. To prevent spurious interactions between periodic slabs, a vacuum of at least 20 Å along the z-direction is included. Furthermore, a dipole correction between periodic slabs is also included [74]. We construct p(2×2) supercells from the primitive unit cells of MXenes and then optimize their lattice constants through a variable cell force minimization. We then calculate the three-dimensional charge density of the force-minimized structures. From this charge density, we determine the surface averaged local potential. The difference between the upper vacuum potential and the Fermi level gives the work function. Since all considered MXenes are symmetric, determining the work function using either the upper or the lower vacuum level yields similar results. An $8 \times 8 \times 1$ Monkhorst-Pack [75] k-point grid is used to sample the Brillouin zone of the p(2×2) supercell. The optimized lattice constants and the work functions of the above mentioned MXenes are listed in table S7 of the supporting information. Graphics of the atomic structures are created using VESTA [76].

2.3. Synthesis, characterisation and work function measurements

The Ti_3AlC_2 MAX phase was purchased from Laizhou Kai Kai Chemical and etched in 48 wt.% hydrofluoric acid (Sigma Aldrich/695068) according to the synthesis process stated in Alhabeb *et al* [77]. The synthesis was prolonged to five days to ensure that the material is fully etched with a saturation of fluorine on the surface, thus yielding Ti_3C_2 MXenes that are terminated with F^*. Ti_3C_2 with nitrogen surface terminations, N^*, was obtained by annealing around 100 mg of F^* terminated Ti_3C_2 under a constant flow (10 sccm) of 10% ammonia in argon (10% NH_3: 90% Ar, Air Products) for 7 h at 700 °C. On the other hand, Ti_3C_2 with oxygen termination, O^*, was obtained by annealing the same amount of F^* terminated Ti_3C_2 under 99.9999% Argon (Leeden National Oxygen Ltd) for 7 h at the same temperature. The samples were characterized with scanning electron microscopy (FESEM JEOL JSM-7600F) (figure S9 in the supporting information), x-ray diffraction (Bruker D8 Advance) (figure 10 in the supporting information), x-ray photo-electron spectroscopy and ultraviolet photoelectron spectroscopy (UPS) (Kratos AXIS Supra). The work function was measured using the following approach. Powder samples of the treated MXenes were dispersed in methanol and drop casted onto electrodeposited gold (<100 μm) on a Si wafer to achieve a thin layer covering the entire surface. The samples were measured via UPS (Kratos Axis Supra) using He I_α

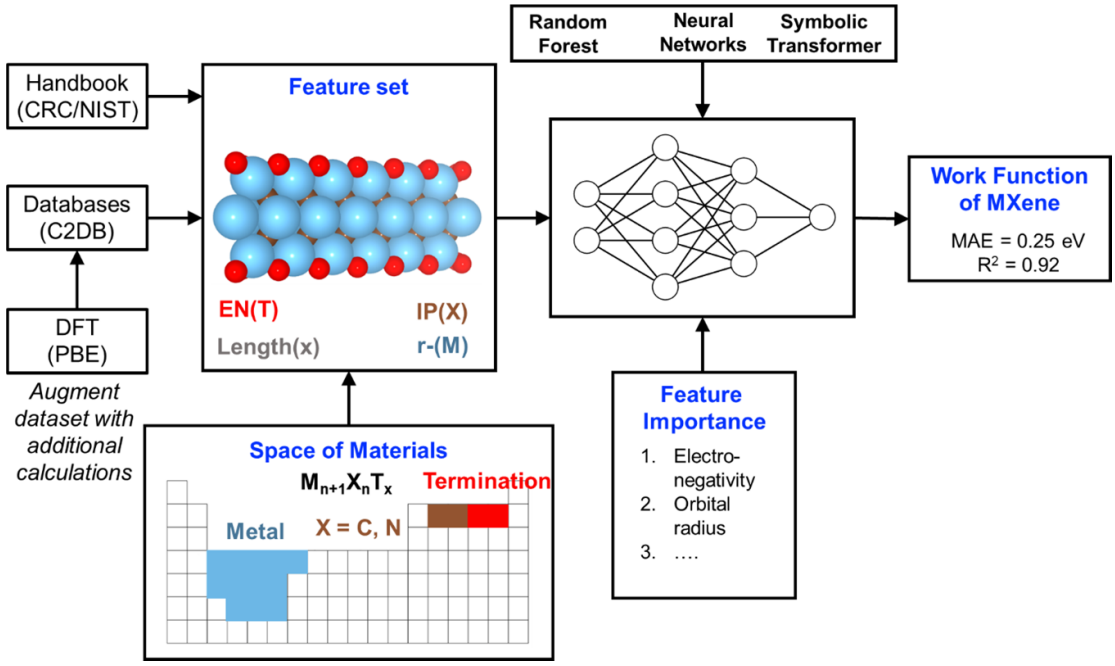

Figure 1. Computational framework for predicting the work function of MXenes (general formula of $M_{n+1}X_nT_x$ where X is either C or N) using machine learning methods. The machine learning methods are trained on work function data from the C2DB. We build a set of 15 features composed of physico-chemical properties like the electronegativity of the termination atoms (EN(T)), ionisation potential of carbon/nitrogen (IP(X)), orbital radius (outermost d-orbital) of the metal atom (r-(M)), and the lattice constant (Length(x)). In addition to training machine learning models to predict the work function, we also identify the most important features that govern trends in the work function across a wide range of MXenes.

($h\nu = 21.22$ eV) as radiation source at the range of −5 eV–20 eV kinetic energy. The work function was extracted from this UPS measurement.

3. Results and discussion

The overall workflow employed is illustrated in figure 1. In section 3.1, we construct a series of machine learning models to predict the work function of MXenes using the 15-feature dataset. After training the machine learning models, in section 3.2, we then rank order the features used to describe the MXenes based on how strongly these features influence the work function, thus adding physical interpretability to our model. Section 3.3 elucidates how reduced-feature datasets can predict the work function without sacrificing on the accuracy of the machine learning models. Finally, in section 3.4, we employ these reduced feature datasets and demonstrate the transferability of our model to new MXene compositions.

3.1. Building machine learning models to predict the work function

To build a machine learning model, the first task is to identify a set of input variables or features that are phenomenologically linked to the target property being predicted. Herein, the target property of interest is the work function. Our feature set consists of 15 features comprising of physico-chemical properties of elements constituting the MXene (e.g. electronegativity), structural properties of the MXene (e.g. lattice constant), and thermodynamic properties reflecting the stability of the MXene (e.g. heat of formation). This set of 15 features is listed in table S2 of the supporting information. Our objective is to train machine learning models that can predict the work function of MXenes having any generic composition using inputs composed of readily available physico-chemical properties of the elements constituting the MXene. We begin with employing ordinary linear regression as a base-case scenario, before pivoting to non-linear models like random forests and neural networks. A five-fold cross validation approach is employed wherein we train the model on any four folds (220 data points) while the fifth fold (55 data points) functions as the test set. Thus, the model is computed five times. The performance of our model is evaluated using the MAE. The ordinary linear regression model containing 15 features, when computed over the five folds, yields an MAE of 0.79 eV on the training data and an MAE of 0.87 eV on the testing data. Further details like the parity plot and the frequency distribution of the residuals are in figure S3 of the supporting information. We did not observe any improvement in the model performance by using either the Lasso or the Ridge regularization techniques. The high MAE on both training and testing data hints that the features are not all linearly independent.

To overcome the shortcomings of the linear regression model, we employ non-linear models like random forest decision trees and neural networks. We evaluate the performance of the random forest model using a five-fold cross validation on the dataset. The random forest model yields an MAE of 0.10 eV on the training data and an MAE of 0.27 eV for the testing data. Both MAEs are averaged over the five folds. The reduced MAE on both the training and testing data in comparison with linear regression hints that the feature set and the work function are non-linearly correlated. In section 3.3, we will use symbolic transformers to further elucidate these non-linear relations. The parity plot and frequency distribution of the residuals for the random forest model is shown in figure S4 of the supporting information.

In addition to the random forest model, we also train a neural network to predict the work function of MXenes. The neural network is trained using a five-fold cross validation approach. The training (176 data points) and validation set (44 data points) accounts for 80% of the total data set. These training and validation data points are used to train the model and optimize the hyperparameters. The remaining 20% of the data set (55 data points) is used for testing the neural network. We thus use a test/train/validation ratio of 55/176/44. Additional details about the hyperparameters employed are listed in table S4 of the supporting information. The parity plots comparing model predictions on the training and test sets for each of the five folds are shown in figure 2. In general, our model behaves similarly across all the five folds. From the fold-specific history plots in figure 2, we observe that the least absolute error difference between the validation and training loss is observed near 600 epochs. Thus, to prevent overfitting, we train the model until 600 epochs. When averaged across all five folds, the model predicts the work function with an MAE of 0.12 eV on the training data and 0.25 eV on the testing data. This MAE of the neural network is comparable to the MAE of the random forest model (0.27 eV).

In figure 3(a), we show a parity plot comparing the model-predicted work function values for MXenes in the test set (over all five folds) to their DFT-derived values from C2DB [52, 53]. We note the predicted values are generally in line with the observed values from the database. To corroborate the findings of the parity plot, we construct a histogram for the residual values as shown in figure 3(b). The residual values represent the difference between model-predicted and database-observed work function values. Based on the histogram, 57% of the MXenes in the test set have a residual value of within $\pm$0.20 eV, indicating the robust performance by our model. The region between $\pm$0.20 eV is specifically selected since it represents the typical accuracy of machine learning models that are trained to reproduce first principles derived energetic properties like adsorption energies of catalytic descriptors, stabilities of nanostructures, and electronic band gaps [57, 78–83].

Despite the model yielding reasonable accuracy across a plurality of materials within the blind test set, one point emerges as an obvious outlier. This outlier is seen in the 15-feature neural network (figures 2 and 3) and reduced-order feature-sets in figures 7, 8, and 10. This outlier is $Sc_4N_3O_2$ with a work function value of 7.23 eV from C2DB. We calculate the work function of $Sc_4N_3O_2$ in this study and obtain a value of 7.18 eV (PBE functional, same as the database). This close agreement indicates that the high error is not due to an incorrect entry in the database. MXenes get oxidized over time [30] and the work function for such O-terminated MXenes is higher than other surface terminations, as seen in figure 5 [32, 33]. To our knowledge, $Sc_4N_3O_2$ has not yet been synthesized, and hence the work function of this MXene cannot be compared with experimental measurements. It appears that the feature set used in this study (table S1 of the supporting information) is unable to accurately estimate the work function for $Sc_4N_3O_2$. This poor estimate could be because the electronic structure of $Sc_4N_3O_2$ warrants a different feature set, with such a study on the electronic structure being part of future work. Our objective is to establish a pre-screening tool that can estimate the work function of MXenes using pre-tabulated properties of constituting elements as inputs, and not to predict the work function accurately for every material. As figure 3 shows, 57% of the MXenes in the test set have predicted values that are within $\pm$0.20 eV from the training set, indicating the broad accuracy shown by our model across a wide-range of MXenes.

3.2. Identifying the properties of MXenes that strongly influence the work function

To build machine learning models trained to reduced-order feature sets, we first need to identify the principal features that govern the work function of MXenes. We thus perform a feature importance analysis on the 15-feature neural network. We use permutation feature importance over the complete dataset to compute individual feature importance values. Permutation feature importance is applied after the model is trained on the entire dataset. The values of a single feature are randomly shuffled, while keeping the other feature values the same. This is followed by making predictions based on this new dataset. These predictions are then compared with the predictions from the initial non-shuffled dataset using a model score (negative MAE). Based on the decrease in the model score, the feature importance is calculated. The higher the decrease in the model score, the higher the importance of the feature. This process is repeated multiple times, and the mean permutation feature importance value is computed. For a specific feature, the feature

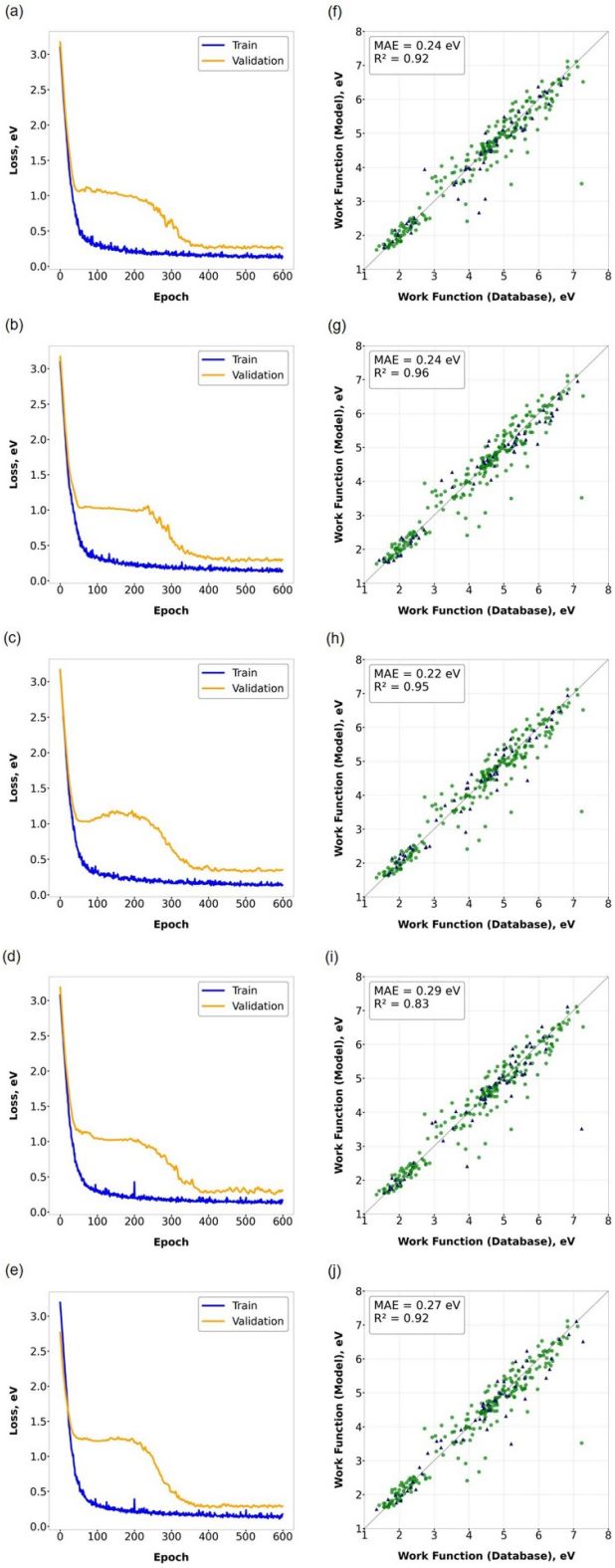

Figure 2. (a)–(e) History plots for each of the five folds in the 15-feature neural network model. The orange graph highlights the variation in the loss (MAE) for validation data, as the number of epochs in the model increase. The blue graph highlights the variation in the loss (MAE) for training data with increasing number of epochs. (f)–(j) Parity plots for each of the five folds of the 15-feature neural network model. These plots compare model predicted work function (eV) with the work function (eV) obtained from the C2DB. Within each fold-specific parity plot, the training set is represented by green circles, while the test set is represented by blue triangles.

importance of the model is computed as the mean permutation feature importance value among five shuffles divided by the sum of all the mean permutation feature importance values. The feature importance values are rank ordered in figure 4. The most important features are termination-specific features such as

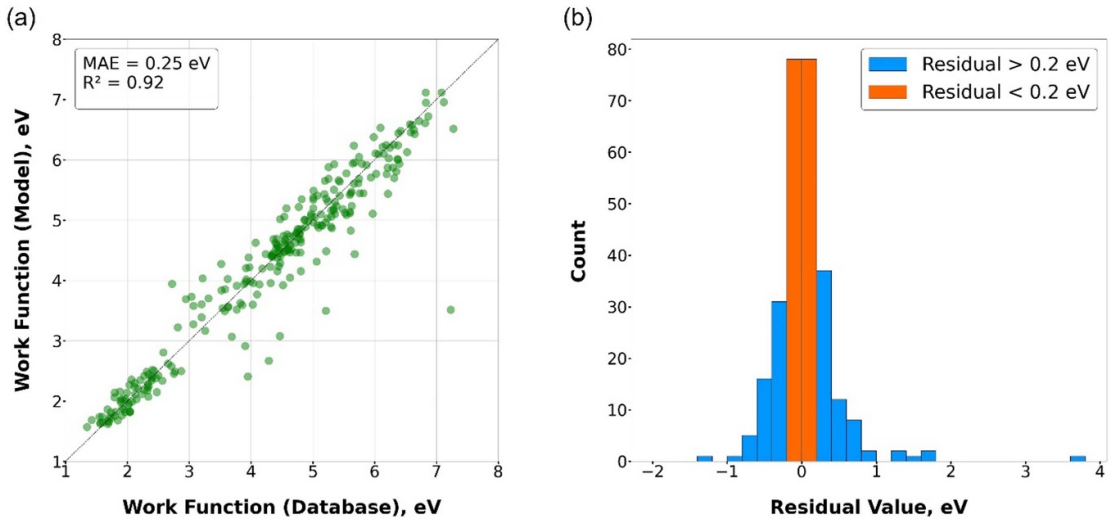

Figure 3. (a) Parity plot comparing the performance of the 15-feature neural network model on the test set. Data from the test sets of all five folds are combined. (b) A histogram representing the residual values (eV) for the test set across all five folds of the 15-feature neural network model. The *y*-axis represents the number of structures corresponding to a given range of residual values along the *x*-axis.

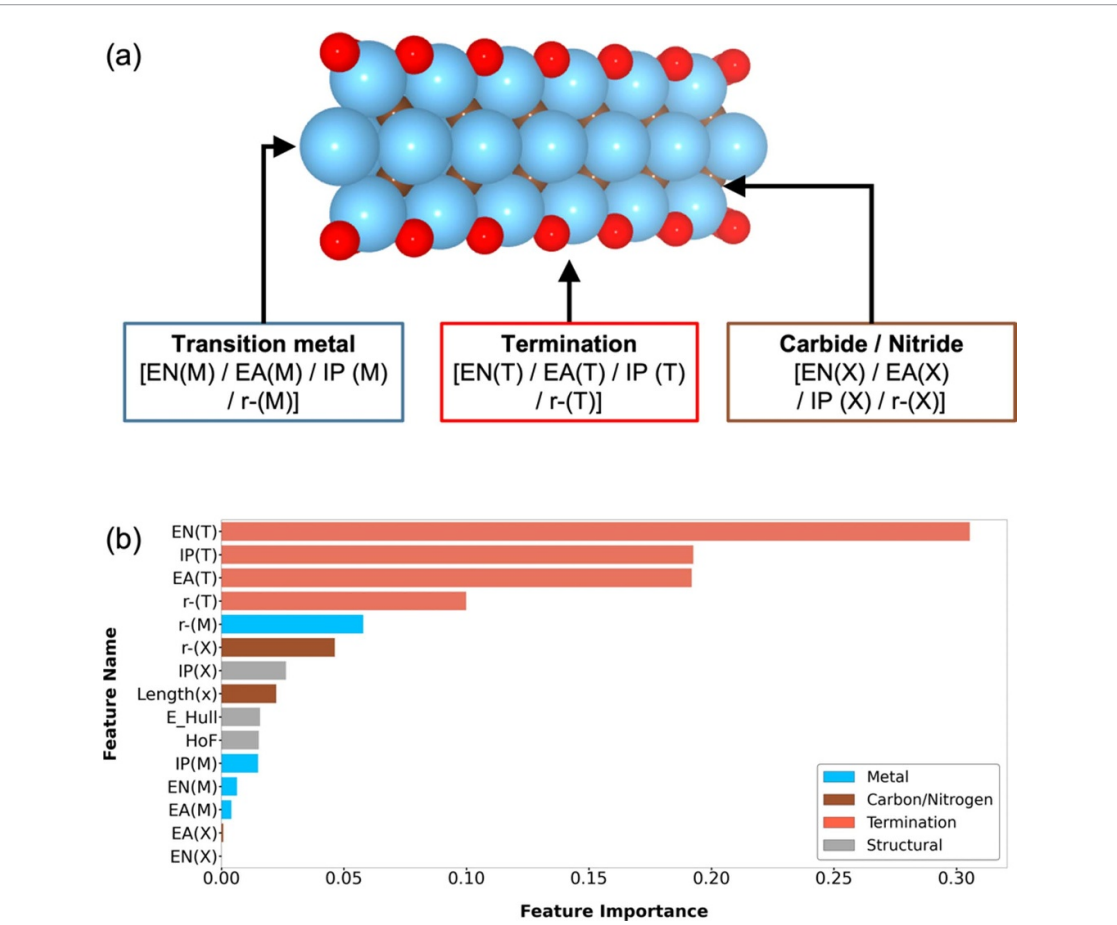

Figure 4. (a) Atomic structure of a prototypical MXene ($M_{n+1}X_nT_x$) with the metal atoms in blue, the carbon/nitrogen atoms in brown, and the atoms terminating the surfaces in red. The features specific to the elements constituting the MXene are listed. EN: electronegativity, EA: electron affinity, IP: ionisation potential, and r-(X), r-(M), r-(T): orbital radii of the outermost orbitals of the M, X, or T atoms. (b) Bar plot showing the importance of each feature in the 15-feature neural network model. As observed, the termination-specific features (red bars) have the highest feature importance.

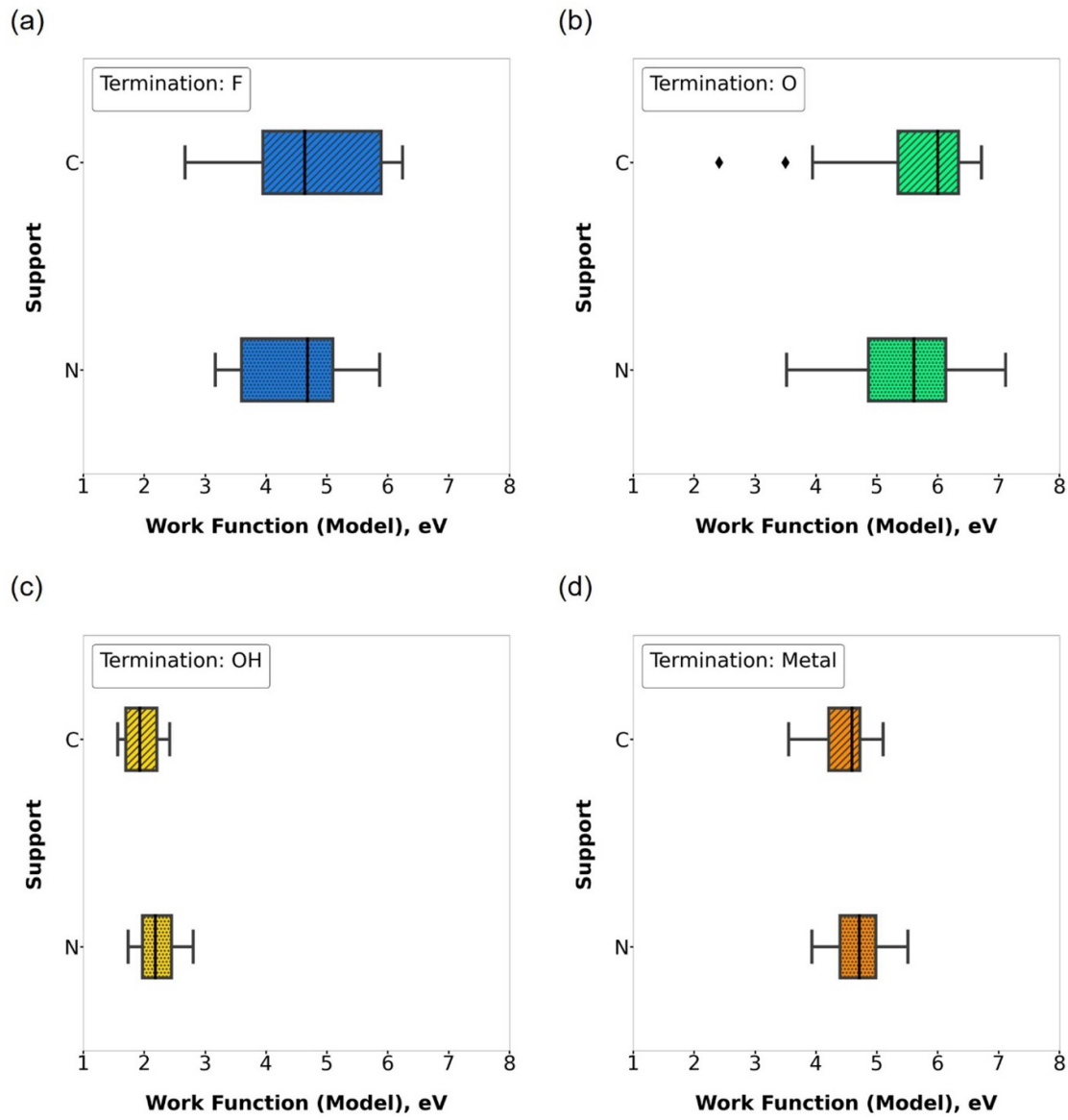

Figure 5. Box plots highlighting how the model predicted work function (15-feature neural network) varies across MXenes that are carbides or nitrides, and for MXenes having different surface terminations. (a) F*-terminated. (b) O*-terminated. (c) OH*-terminated. (d) Bare-metal-terminated. Boxes with the slant and dotted pattern represent the transition metal carbides and nitrides respectively. The overall trends indicate that changing the surface termination of the MXene strongly influences the work function. On the other hand, for a given surface termination,changing the MXene from a carbide to a nitride has a relatively smaller influence on the work function.

electronegativity (EN(T)), electron affinity (EA(T)), ionisation potential (IP(T)), and radius of orbitals of the termination atoms (r-(T)). Metal-specific features such as radius of orbitals of the metal atoms (r-(M)) are also dominant. This feature importance study reveals that the value of the work function is dictated by the properties of the surface termination, with the electronegativity of the termination being the most important property (feature importance of 0.31). In section 3.3, we will compare the insights obtained through permutation feature importance with two additional complementary approaches that gauge feature importance. Taken together, these studies lend physical interpretability to our analysis by identifying the set of features most strongly influencing the work function of MXenes.

To confirm the dominance of termination-based features in determining the work function, we construct box plots in figure 5. These plots categorize the model predicted work function of MXenes ($M_{n+1}X_nT_x$) in terms of the surface terminations ($T_x = OH^*$, F^*, O^*, and bare metal), and whether the MXenes are a carbide (X = C) or a nitride (X = N). Although qualitative, the trends in the work function are more sensitive to the nature of the surface termination, than whether the MXenes are carbides or nitrides. Furthermore, OH^* terminated MXenes have the lowest work function as has been observed in previous studies [31]. On the other hand, O^* terminated MXenes have the highest work function. Another general trend observed is that the work function increases as the surface termination is altered from F^*

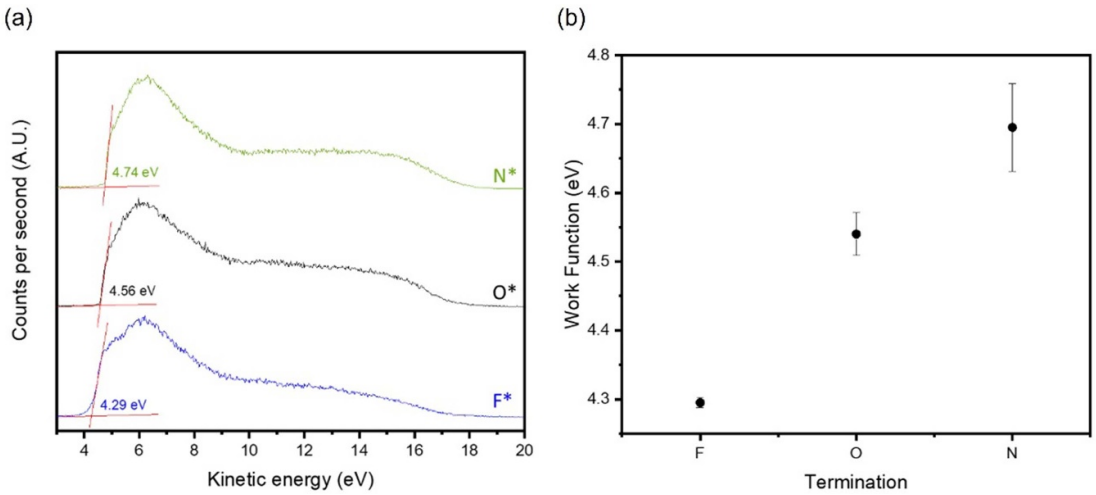

Figure 6. (a) UPS measurement and extraction of the work function from the kinetic energy for Ti_3C_2 MXenes that are terminated with N*, O*, and F*. (b) Trends in the extracted work functions of $Ti_3C_2T_x$ with different terminations; F*, O*, and N*. The effect of the termination is observed to be significant.

(as-synthesized MXenes) to O*. The importance of the surface termination in determining the work function is consistent with observations from the feature importance analysis discussed above.

Based on insights from the machine learning model, we performed an experimental study to understand how the surface termination influences the work function. We synthesized MXenes based on the guidelines in Alhabeb *et al* [77], but with prolonged etching to saturate the fluorine termination (termed as F*). Subsequently, we modified the termination using simple gaseous annealing techniques to induce a change in the termination from F* in the as-synthesized case, to O* (annealing in argon) and N* (annealing in ammonia). The work function for the resultant materials is measured using UPS and extracted from the onset of the UPS spectrum [84] as depicted in figure 6(a). Trends in the work function across different terminations are shown in figure 6(b). The F* terminated Ti_3C_2 has a work function of around 4.3 eV, which is close to the ML model predicted work function value of 5 eV. Upon annealing in argon, which induces an oxygen termination, the work function increases by 0.2 eV–4.5 eV. Our machine learning model indicates that changing the termination from F* to O* also increases the work function to 6.1 eV, though the magnitude of this increase is larger than experiments. Possible reasons for the discrepancy include the formation of TiO_x species on the O* terminated Ti_3C_2. Annealing in ammonia induces a nitrogen termination N*, which results in an even higher experimentally measured work function of 4.7 eV. Despite the limited terminations that can be achieved experimentally, these experiments demonstrate that the surface termination readily influences the work function.

3.3. Predicting the work function using reduced feature datasets

Although the 15-feature dataset yields relatively accurate machine learning models, the Pearson correlation coefficients (see figure S1 in the supporting information) indicate that some of the features are highly correlated. Hence, we anticipate the possibility of training simpler machine learning models that contain fewer features in the dataset. Such reduced-feature datasets are more readily transferable to MXenes of different compositions. In this section, we systematically reduce the 15-feature datasets and show that the resulting models have comparable accuracy. In the first instance, we discard all features that have a Pearson correlation above 75% (see table S2 and figure S2 in the supporting information) thus creating a ten-feature dataset. These ten features include two DFT-derived energetic properties; the heat of formation, and the energy above the convex hull. We then create an eight-feature dataset that excludes these two DFT-derived properties (see table S2 in the supporting information). The resulting eight-feature dataset is more readily transferable to new MXene compositions whose heat of formation and the energy above the hull may not have been computed. Finally, we discuss machine learning models that are trained to a more minimalistic five-feature dataset.

The ten-feature dataset is used to train the two best-performing machine learning models; the random forest and the neural network. For the ten-feature random forest regression model, the work function is predicted with an MAE of 0.11 eV on the training data and 0.32 eV on the testing data, when averaged over five-folds (see figure S5 in the supporting information). On similar lines, the ten-feature neural network model predicts the work function with an MAE of 0.13 eV on the training data and 0.26 eV on the testing

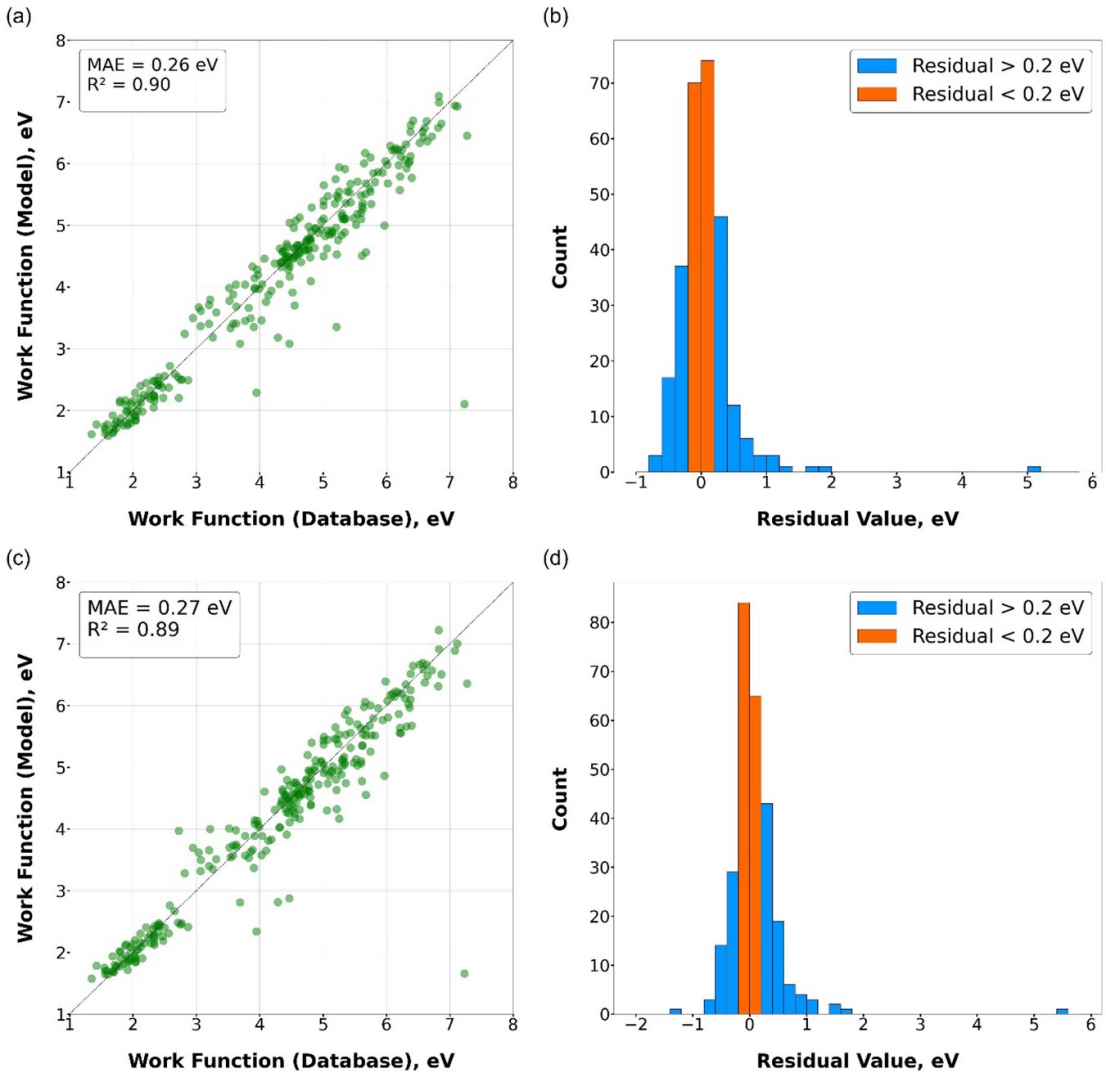

Figure 7. (a) Parity plot comparing the performance of the ten-feature neural network model on the test set. Data from the test sets of all five folds are combined. (b) A histogram representing residual values (eV) for the ten-feature neural network model. (c) Parity plot comparing the performance of the eight-feature neural network model on the test set. Data from the test sets of all five folds are combined. (d) A histogram representing residual values (eV) for the eight-feature neural network model. The *y*-axis represents the number of structures corresponding to a given range of residual values along the *x*-axis.

data, when averaged over five-folds. From the parity plot in figure 7(a), we observe that the model predicted work function values are in line with the work function from C2DB. The model accuracy is further confirmed through the histogram plot of the residual values in figure 7(b).

We observe that the highest feature importance is exhibited by termination-specific features such as electronegativity (EN(T)) and the radius of orbital of the termination (r-(T)) (figure S6 in the supporting information). Despite using a ten-feature dataset, we find that the relative importance of features remains unchanged with the termination-specific features being dominant. More crucially, the 10-feature machine learning model (random forest and neural network) yields nearly the same performance as the 15-feature machine learning model.

To obtain an even simpler and less computationally expensive model, we construct an eight-feature dataset. This dataset is constructed by removing two DFT-derived features from the ten-feature dataset, the heat of formation, and the energy above the convex hull. The eight-feature dataset is trained on the neural network model and predicts the work function with an MAE of 0.14 eV on the training data and 0.27 eV on the testing data (averaged over five-folds). The model performance is similar to the 15-feature and 10-feature neural network models, and the accuracy is largely unaffected by the removal of the two DFT-derived features. The model performance is depicted in the parity plot in figure 7(c), and the histogram plot of the residual values in figure 7(d).

Lamoureux *et al* [61] showed the power of creating minimalistic machine learning models while attempting to predict nanoparticle stability from the properties of the constituting elements. These

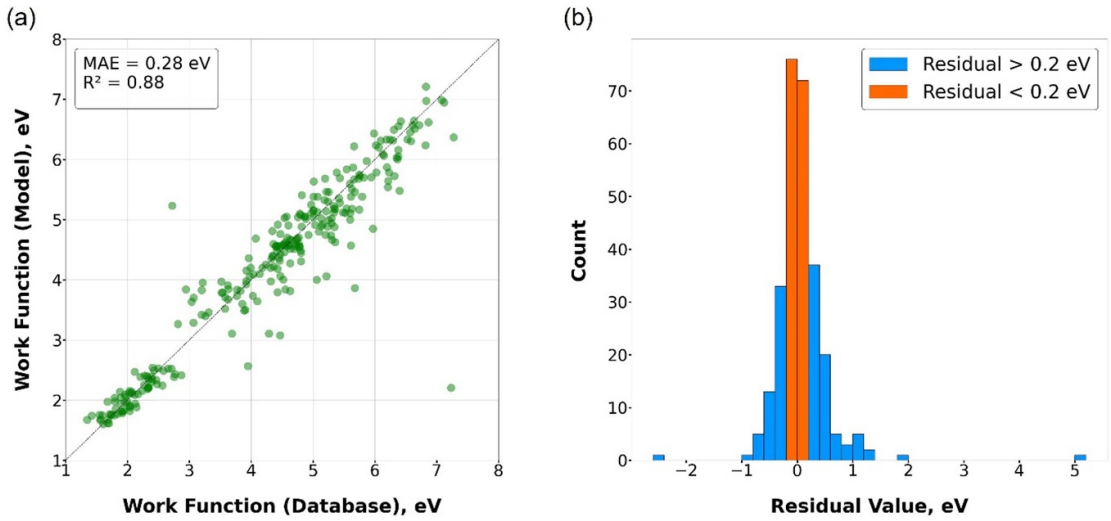

Figure 8. (a) Parity plot comparing the performance of the best performing five-feature neural network model on the test set. Data from the test sets of all five folds are combined. (b) A histogram representing residual values (eV) for this neural network model. The *y*-axis represents the number of structures corresponding to a given range of residual values along the *x*-axis.

minimalistic models were not only more transferable, but also retained the most important features. Such models can thus independently isolate the most crucial features thereby complementing the feature importance analysis shown in figure 4. Taking the ten-feature dataset as a reference, we construct all possible combinations of five-feature datasets. The rationale for limiting our analysis to five-feature datasets is discussed further below. This process yields 252 different combinations of features. The neural network machine learning model is trained to all 252 combinations of different five-feature datasets and the resulting models are rank ordered based on increasing MAE in the work function prediction. The MAE is the average value after performing a five-fold cross validation. The best performing five-feature neural network employs the following features: (a) electronegativity of the termination (EN(T)), (b) radius of orbital of the termination (r-(T)), (c) radius of orbital of the metal (r-(M)), (d) ionisation potential of the support (IP(X)) and (e) the lattice constant (L(x)). This model predicts the work function with an MAE of 0.16 eV on the training data and 0.28 eV on the testing data, when averaged over all five folds. From the parity plot in figure 8(a) and the histogram plot of the residuals in figure 8(b), we observe most model-predicted values are consistent with the database-observed values. Thus, as the number of features is progressively decreased from $15 \rightarrow 10 \rightarrow 8 \rightarrow 5$, the model accuracy remains relatively unchanged. Moreover, the five features that compose the best performing neural network are broadly consistent with the feature importance analysis of the 15-feature model shown in figure 4. As we further reduce the feature space from five to four, we note that the MAE increases slightly, from 0.28 eV (testing data) to 0.30 eV (testing data). Hence, we limit ourselves to using the five-feature model as a minimalistic model for predicting the work function.

For our subsequent analysis, we consider the top 10% of all models (25 models) from the 252 combinations. To understand trends in the different features within these 25 models, we introduce a metric called the occurrence probability. The occurrence probability is the ratio of the frequency of a feature within the top-25 models, to the total frequency of all the features within these top 25 models. Thus, the maximum occurrence probability for a feature is 25/125, if this feature were to occur in all the top-25 five-feature models. A sample calculation of this metric is shown in figure 9(a). This occurrence probability is an alternative approach of gauging feature importance, since the metric reflects the frequency of a given feature in the top-25 models. The electronegativity of the termination (EN(T)) and the lattice constant (L(x)) display the highest occurrence probability, followed by the radius of orbital of the metal (r-(M)) as seen in figure 9(b). This analysis corroborates the observation that trends in the work function of MXenes are dictated by termination-specific features with the electronegativity of the termination being the most crucial input. These observations are wholly consistent with the feature importance studies for the 8-feature (figure S7 in the supporting information), 10-feature (figure S6 in the supporting information) and 15-feature models (figure 4).

Another approach to identify the most important features within a machine learning model is to use a genetic algorithm for feature selection [61, 85]. To that end, we employ a symbolic transformer via genetic programming to construct complex metal-features that are strongly correlated with the work function of MXenes. Through this analysis, we aim to independently support the importance of termination-based

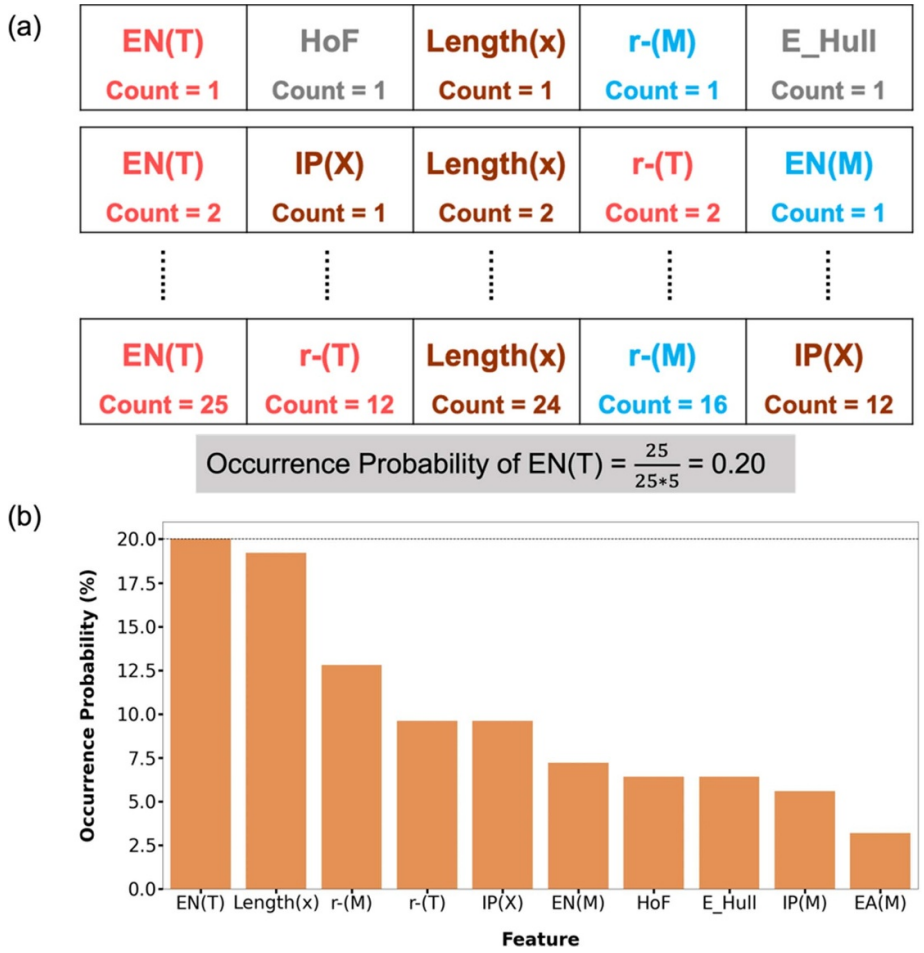

Figure 9. (a) The approach to calculate the occurrence probability of a specific feature within the top-25 models. The occurrence probability is the frequency of a feature used in the top-25 models divided by the total frequency of all the features among the top-25 models. (b) A bar plot illustrating the occurrence probability of each feature in the top-25 five-feature neural network models. Consistent with the feature importance analysis in figure 4, we find that the termination-specific features occur with the highest frequency.

features in determining the work function, as evinced from the feature importance (figure 4) and occurrence probability (figure 9) analysis. We clarify that the purpose of the genetic programme is not to bring forth the most readily interpretable mathematical forms, but to isolate the identity of key features that most strongly correlate with the work function.

We use a symbolic transformer within a genetic programme, fitted on the eight-feature data set. The symbolic transformer constructs meta-features that are composed of different three-feature combinations connected with mathematical operators (table S5 of the supporting information). The fitness metric used to evaluate these meta-features is the Pearson correlation coefficient with the work function (target variable). The genetic programme constructs the fittest meta-features that possess the strongest Pearson correlation coefficient with the work function. Those meta-features with fitness above 90% are collected separately. The top-ten meta-features with the highest fitness are listed in table S6 of the supporting information. We note that all the meta-features with a fitness above 90% primarily consist of features representing the surface termination of MXenes (EN(T) and r-(T)). Thus, the evolutionary algorithm confirms the dominance of the termination-related features in determining the work function, further corroborating our machine learning models. We acknowledge that the meta-features most strongly correlating with the work function in table S6 of the supporting information are not readily interpretable. In principle, the meta-features should provide clues regarding the underlying physical phenomena. In this case however, the genetic programme identifies the primary features (EN(T) and r-(T)) governing the work function from the entire feature-set used (table S1 of the supporting information. Hence, the scope of using the genetic programme is restricted to independently corroborating feature importance. Indeed, the dominant role of termination-based features, especially the electronegativity is consistent with other methods employed like permutation feature importance (figure 4) and occurrence probability (figure 9). Upon confirming key attributes of our machine

learning models using reduced feature datasets, we will now discuss the transferability of our models to new MXene compositions beyond the original dataset.

3.4. Transferability of machine learning models to MXenes beyond the dataset

The previous sections outlined a series of machine learning models that predict the work function of MXenes having different surface terminations like F*, O*, OH*, and bare metal atoms. Furthermore, the physico-chemical properties of the surface terminations emerged as the most important features governing the work function. The range of possible surface terminations of MXenes has recently been expanded to include Br*, Cl*, S*, NH*, and N*. The ultimate test for our model will be in predicting the work function across a wider range of surface terminations that go beyond the training set. Previous studies [86] have shown that transfer learning is an effective way towards extending the validity of a model over a wider range of materials. In transfer learning, only user specified layers of the neural network are retrained while the weights corresponding to other layers remain unchanged. Herein, we evaluate the transferability of our eight-feature dataset (shown in figure 7) using transfer learning. Using DFT, we compute the lattice constants and the work function of 40 additional MXenes having Br*, Cl*, S*, NH*, and N* terminations on the surface. These values are listed in table S7 of the supporting information. The features for these 40 additional MXenes can be directly evaluated from databases and do not require additional DFT calculations. In transfer learning, we first train the eight-feature dataset on the 275 MXenes composed of F*, O*, OH*, and bare metal atoms as surface terminations. We then disable the training for all layers except the final hidden layer and the output layer. This modification ensures that only the parameters of the final hidden layer and the output layer will be updated, as the 40 additional MXenes having Br*, Cl*, S*, NH*, and N* terminations are added to the dataset.

The pre-trained weights from the eight-feature neural network model (trained to 275 data points) model are used in the transfer learning neural network model, on a secondary dataset of the 40 MXenes that is described above. The MXenes in the secondary dataset are split randomly such that each fold contains at least one MXene of each surface termination. The work function is predicted with an MAE of 0.58 eV on the training data and 0.76 eV on the testing data (averaged over five-folds). The high MAE indicates that transfer learning models are not sufficiently accurate when tasked with predicting the work function of MXenes having surface terminations that are different from the original training set. This high error on the new set of surface terminations that are unseen by all the weights of the neural network, is not entirely unexpected. This is because the work function strongly depends on physico-chemical properties of the atoms terminating the MXene surface.

We now re-train the eight-feature neural network model on an augmented dataset of 315 MXenes. This augmented dataset consists of 275 structures from C2DB (F*, O*, OH*, and bare metal surface terminations), and 40 structures from the new set of terminations (Br*, Cl*, S*, NH*, and N*). Five-fold cross validation is used to train the neural network. The 40 additional MXenes are split equally among the five folds of the original dataset (55 MXenes from C2DB per fold) such that each fold contains at least one MXene of each termination. This leads to each fold containing 63 MXenes in total, 55 MXenes from C2DB and 8 MXenes from the new set of surface terminations. The augmented eight-feature neural network model has a test/train/validation ratio of 63/202/50. The MAEs on the training and test set, averaged over five folds, are 0.16 eV and 0.28 eV, respectively. The parity plot is shown in figure 10. This example illustrates how machine learning models trained on data from open-access databases, that is supplemented by a limited set of DFT calculations, can expand the materials space across which properties like the work function are predicted.

This work shows that the work function is principally dependent on the properties of the functional group terminating the MXene. The surface dipoles formed in the terminated surface govern the work function value [31]. Our study considers the simple situation of a single surface termination (e.g. either O*, F*, OH*, or bare metal) being present. However, MXenes with mixed terminations are routinely synthesized and are shown to be stable in computational studies. The distribution and mixing energies of MXenes terminated with more than one element are more strongly influenced by interactions among surface termination groups than on the metal type, carbide/nitride, or the MXenes thickness [32]. We perform a computer experiment to verify the importance of the connectivity of the MXene structure in determining the work function of mixed-terminated MXenes. Our feature set (table S1 of the supporting information) does not include such connectivity-based information. If the connectivity-based information of mixed-terminated MXenes is important, then our feature set will be unable to predict the work function of MXenes containing mixed terminations. Using DFT calculations, we calculate the work function of $Ti_3C_2O_{0.44}N_{0.56}$ across a series of different local arrangements of O* and N*. We find that the work function strongly depends on the local arrangement of O* and N*, with variations of 0.36 eV across the different configurations tested, as illustrated in figure S9. Hence, connectivity-based information is indeed important

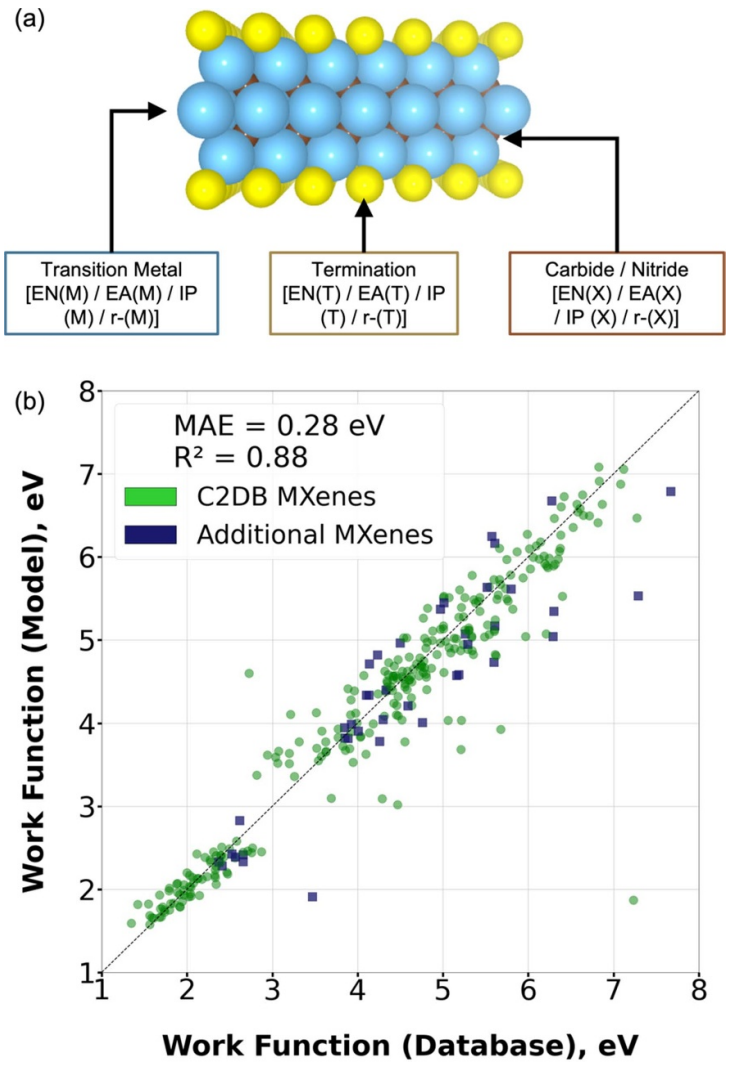

Figure 10. (a) Atomic structure of a generic MXene along with the features like physico-chemical properties of elements that are obtained directly from handbooks and databases. The transition metal, carbon/nitrogen, and surface terminations are indicated in blue, brown, and yellow, respectively. (b) Parity plot comparing the performance of the eight-feature neural network model on the test set. Data from the test sets of all five folds are combined. Green circles indicate model predictions for the 275 MXenes within C2DB as compared with the database values. Blue squares compare model predictions on the 40 additional MXenes having difference surface terminations like Br*, Cl*, S*, N*, and NH* with DFT-derived work function values.

in predicting the work function of mixed-terminated MXenes. For mixed-terminated MXenes, features obtained from the composition of the MXene need to be augmented by a graph-based connectivity model [46, 87]. Extending our model to MXenes with mixed terminations will be part of future work.

4. Conclusions

The work function of MXenes is an important fundamental property governing their applications in catalysis, energy storage and conversion, and in electronic devices. Herein, we present a series of machine learning models that predict the work function of MXenes obtained from the C2DB. Our models are extendable to MXenes having generic compositions and a wide range of surface terminations (O*, F*, OH*, and bare metal atoms). Our machine learning models use a set of 15 features that are broadly classified into: the structural properties of MXenes (lattice constants); properties of elements constituting MXenes (ionisation potential, electron affinity, electronegativity, and orbital radius); and the thermodynamic properties of MXenes (energy above hull and the heat of formation). Both the 15-feature random forest model and the 15-feature neural network model predict the work function of MXenes with an MAE of 0.27 eV and 0.25 eV on the test set, respectively. We construct a series of reduced-order models containing 10-, 8-, and 5-features that predict the work function of MXenes with accuracies on par with the 15-feature model. These reduced-order models also aid in isolating the most important features governing the work function. Both the reduced-order models and a feature importance analysis indicate that the work function of MXenes is most strongly

influenced by properties of the surface termination. We show that the electronegativity of atoms terminating the MXene surface emerges as the most important feature. These insights are confirmed through experiments illustrating the sensitivity of the work function as the surface termination of Ti_3C_2 is changed from F* to O* to N*. We finally showcase the transferability of our eight-feature neural network to 40 MXenes with different surface terminations (Br*, Cl*, S*, N*, and NH*) than those taken from the C2DB. Retraining the 8-feature neural network on this augmented dataset results in the work function being predicted with an MAE of 0.28 eV, an accuracy comparable to that of the original 15-feature neural network. This example illustrates that the materials space for which a given property is predicted can be widened by training models on open-access databases that are supplemented with a limited set of new DFT calculations. Taken together, these machine learning models will enable the rapid identification of MXenes having work function values within a targeted range, for their applications in catalysis, energy storage, and electronics.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This work is supported by the National Research Foundation (NRF), Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) program and from the Ministry of Education Academic Research Fund Tier 1: RS04/19 and RG 5/22. We thank the High-Performance Computing (HPC) team at the HPC Centre, Nanyang Technological University (NTU) for technical assistance and computing resources. The computational work for this article was partially performed on resources of the National Supercomputing Centre, Singapore (www.nscc.sg) through Project IDs 12001868, 12002171 and 12002494. P R acknowledges the Visiting Research Student Programme by the India Connect@NTU office. L R acknowledges NTU for a research scholarship. We acknowledge Professor Kristian Thygesen from the Technical University of Denmark for sharing the C2DB file with us.

ORCID iDs

Lavie Rekhi https://orcid.org/0000-0002-1039-9292
Tej S Choksi https://orcid.org/0000-0002-9520-019X

References

[1] Gogotsi Y and Huang Q 2021 MXenes: two-dimensional building blocks for future materials and devices *ACS Nano* **15** 5775–80
[2] Anasori B, Lukatskaya M R and Gogotsi Y 2017 2D metal carbides and nitrides (MXenes) for energy storage *Nat. Rev. Mater.* **2** 16098
[3] Kamysbayev V, Filatov Alexander S, Hu H, Rui X, Lagunas F, Wang D, Klie R F and Talapin D V 2020 Covalent surface modifications and superconductivity of two-dimensional metal carbide MXenes *Science* **369** 979–83
[4] Björk J and Rosen J 2021 Functionalizing MXenes by tailoring surface terminations in different chemical environments *Chem. Mater.* **33** 9108–18
[5] Urbankowski P, Anasori B, Makaryan T, Er D, Kota S, Walsh P L, Zhao M, Shenoy V B, Barsoum M W and Gogotsi Y 2016 Synthesis of two-dimensional titanium nitride Ti_4N_3 (MXene) *Nanoscale* **8** 11385–91
[6] Zhan C, Sun W, Xie Y, Jiang D-E and Kent P R C 2019 Computational discovery and design of MXenes for energy applications: status, successes, and opportunities *ACS Appl. Mater. Interfaces* **11** 24885–905
[7] Li Z *et al* 2019 *In situ* formed Pt3Ti nanoparticles on a two-dimensional transition metal carbide (MXene) used as efficient catalysts for hydrogen evolution reactions *Nano Lett.* **19** 5102–8
[8] Thakur R, VahidMohammadi A, Smith J, Hoffman M, Moncada J, Beidaghi M and Carrero C A 2020 Insights into the genesis of a selective and coke-resistant MXene-based catalyst for the dry reforming of methane *ACS Catal.* **10** 5124–34
[9] Seh Z W, Fredrickson K D, Anasori B, Kibsgaard J, Strickler A L, Lukatskaya M R, Gogotsi Y, Jaramillo T F and Vojvodic A 2016 Two-dimensional molybdenum carbide (MXene) as an efficient electrocatalyst for hydrogen evolution *ACS Energy Lett.* **1** 589–94
[10] Johnson L R *et al* 2020 MXene materials for the electrochemical nitrogen reduction—functionalized or not? *ACS Catal.* **10** 253–64
[11] Khazaei M, Ranjbar A, Arai M, Sasaki T and Yunoki S 2017 Electronic properties and applications of MXenes: a theoretical review *J. Mater. Chem.* C **5** 2488–503
[12] Kim H, Wang Z and Alshareef H N 2019 MXetronics: electronic and photonic applications of MXenes *Nano Energy* **60** 179–97
[13] Li X, Yin X, Liang S, Li M, Cheng L and Zhang L 2019 2D carbide MXene Ti_2CT_X as a novel high-performance electromagnetic interference shielding material *Carbon* **146** 210–7
[14] Iqbal A, Kwon J, Kim M K and Koo C M 2021 MXenes for electromagnetic interference shielding: experimental and theoretical perspectives *Mater. Today Adv.* **9** 100124
[15] Ahn S, Han T H, Maleski K, Song J, Kim Y H, Park M H, Zhou H, Yoo S, Gogotsi Y and Lee T-W 2020 A 2D titanium carbide MXene flexible electrode for high-efficiency light-emitting diodes *Adv. Mater.* **32** e2000919
[16] Chen J, Liu X, Li Z, Cao F, Lu X and Fang X 2022 Work-function-tunable MXenes electrodes to optimize p-$CsCu_2I_3$/n-$Ca_2Nb_{3-x}TaxO_{10}$ junction photodetectors for image sensing and logic electronics *Adv. Funct. Mater.* **32** 2201066
[17] Liu Y, Xiao H and Goddard W A 2016 Schottky-barrier-free contacts with two-dimensional semiconductors by surface-engineered MXenes *J. Am. Chem. Soc.* **138** 15853–6

[18] Wang Z, Kim H and Alshareef H N 2018 Oxide thin-film electronics using all-MXene electrical contacts *Adv. Mater.* **30** 1706656
[19] Yuan Y, Li H, Wang L, Zhang L, Shi D, Hong Y and Sun J 2019 Achieving highly efficient catalysts for hydrogen evolution reaction by electronic state modification of platinum on versatile $Ti_3C_2T_x$ (MXene) *ACS Sustain. Chem. Eng.* **7** 4266–73
[20] Xu C, Fan C, Zhang X, Chen H, Liu X, Fu Z, Wang R, Hong T and Cheng J 2020 MXene ($Ti_0C_2T_x$) and carbon nanotube hybrid-supported platinum catalysts for the high-performance oxygen reduction reaction in PEMFC *ACS Appl. Mater. Interfaces* **12** 19539–46
[21] Cui C, Cheng R, Zhang H, Zhang C, Ma Y, Shi C, Fan B, Wang H and Wang X 2020 Ultrastable MXene@Pt/SWCNTs' nanocatalysts for hydrogen evolution reaction *Adv. Funct. Mater.* **30** 2000693
[22] Zhou H *et al* 2021 Engineering the Cu/Mo_2CT_x (MXene) interface to drive CO_2 hydrogenation to methanol *Nat. Catal.* **4** 860–71
[23] Li Z *et al* 2021 Direct methane activation by atomically thin platinum nanolayers on two-dimensional metal carbides *Nat. Catal.* **4** 882–91
[24] Li Z *et al* 2018 Reactive metal–support interactions at moderate temperature in two-dimensional niobium-carbide-supported platinum catalysts *Nat. Catal.* **1** 349–55
[25] O'Connor N J, Jonayat A S M, Janik M J and Senftle T P 2018 Interaction trends between single metal atoms and oxide supports identified with density functional theory and statistical learning *Nat. Catal.* **1** 531–9
[26] Liu C-Y, Zhang S, Martinez D, Li M and Senftle T P 2020 Using statistical learning to predict interactions between single metal atoms and modified MgO(100) supports *npj Comput. Mater.* **6** 102
[27] Choksi T, Majumdar P and Greeley J P 2018 Electrostatic origins of linear scaling relationships at bifunctional metal/oxide interfaces: a case study of Au nanoparticles on doped MgO substrates *Angew. Chem., Int. Ed.* **57** 15410–4
[28] Yang Y, Zhou J, Nakayama M, Nie L, Liu P and White M G 2014 Surface dipoles and electron transfer at the metal oxide–metal interface: a 2PPE study of size-selected metal oxide clusters supported on Cu(111) *J. Phys. Chem.* C **118** 13697–706
[29] Schultz T, Frey N C, Hantanasirisakul K, Park S, May S J, Shenoy V B, Gogotsi Y and Koch N 2019 Surface termination dependent work function and electronic properties of $Ti_3C_2T_x$ MXene *Chem. Mater.* **31** 6590–7
[30] Halim J, Cook K M, Naguib M, Eklund P, Gogotsi Y, Rosen J and Barsoum M W 2016 X-ray photoelectron spectroscopy of select multi-layered transition metal carbides (MXenes) *Appl. Surf. Sci.* **362** 406–17
[31] Khazaei M, Arai M, Sasaki T, Ranjbar A, Liang Y and Yunoki S 2015 OH-terminated two-dimensional transition metal carbides and nitrides as ultralow work function materials *Phys. Rev.* B **92** 075411
[32] Ibragimova R, Puska M J and Komsa H-P 2019 pH-dependent distribution of functional groups on titanium-based MXenes *ACS Nano* **13** 9171–81
[33] Ibragimova R, Erhart P, Rinke P and Komsa H-P 2021 Surface functionalization of 2D MXenes: trends in distribution, composition, and electronic properties *J. Phys. Chem. Lett.* **12** 2377–84
[34] Kirklin S, Saal J E, Meredig B, Thompson A, Doak J W, Aykol M, Rühl S and Wolverton C 2015 The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies *npj Comput. Mater.* **1** 15010
[35] Chanussot L *et al* 2021 Open Catalyst 2020 (OC20) dataset and community challenges *ACS Catal.* **11** 6059–72
[36] Winther K T, Hoffmann M J, Boes J R, Mamun O, Bajdich M and Bligaard T 2019 Catalysis-Hub.org, an open electronic structure database for surface reactions *Sci. Data* **6** 75
[37] Andersen M, Levchenko S V, Scheffler M and Reuter K 2019 Beyond scaling relations for the description of catalytic materials *ACS Catal.* **9** 2752–9
[38] Li Z, Achenie L E K and Xin H 2020 An adaptive machine learning strategy for accelerating discovery of perovskite electrocatalysts *ACS Catal.* **10** 4377–84
[39] Ma X, Li Z, Achenie L E K and Xin H 2015 Machine-learning-augmented chemisorption model for CO_2 electroreduction catalyst screening *J. Phys. Chem. Lett.* **6** 3528–33
[40] Wang S-H, Pillai H S, Wang S, Achenie L E K and Xin H 2021 Infusing theory into deep learning for interpretable reactivity prediction *Nat. Commun.* **12** 5288
[41] Flores R A, Paolucci C, Winther K T, Jain A, Torres J A G, Aykol M, Montoya J, Nørskov J K, Bajdich M and Bligaard T 2020 Active learning accelerated discovery of stable iridium oxide polymorphs for the oxygen evolution reaction *Chem. Mater.* **32** 5854–63
[42] Tan K, Dixit M, Dean J and Mpourmpakis G 2019 Predicting metal–support interactions in oxide-supported single-atom catalysts *Ind. Eng. Chem. Res.* **58** 20236–46
[43] Wexler R B, Gautam G S, Stechel E B and Carter E A 2021 Factors governing oxygen vacancy formation in oxide perovskites *J. Am. Chem. Soc.* **143** 13212–27
[44] Schlexer Lamoureux P, Winther K T, Garrido-Torres J A, Streibel V, Zhao M, Bajdich M, Abild-Pedersen F and Bligaard T 2019 Machine learning for computational heterogeneous catalysis *ChemCatChem* **11** 3581–601
[45] Jain A, Wang Z and Nørskov J K 2019 Stable two-dimensional materials for oxygen reduction and oxygen evolution reactions *ACS Energy Lett.* **4** 1410–1
[46] Back S, Yoon J, Tian N, Zhong W, Tran K and Ulissi Z W 2019 Convolutional neural network of atomic surface structures to predict binding energies for high-throughput screening of catalysts *J. Phys. Chem. Lett.* **10** 4401–8
[47] Back S, Na J, Tran K and Ulissi Z W 2020 In silico discovery of active, stable, CO-tolerant and cost-effective electrocatalysts for hydrogen evolution and oxidation *Phys. Chem. Chem. Phys.* **22** 19454–8
[48] Back S, Na J and Ulissi Z W 2021 Efficient discovery of active, selective, and stable catalysts for electrochemical H_2O_2 synthesis through active motif screening *ACS Catal.* **11** 2483–91
[49] Chang C and Medford A J 2021 Application of density functional tight binding and machine learning to evaluate the stability of biomass intermediates on the Rh(111) surface *J. Phys. Chem.* C **125** 18210–6
[50] Li J, Halldin Stenlid J, Tang M T, Peng H-J and Abild-Pedersen F 2022 Screening binary alloys for electrochemical CO_2 reduction towards multi-carbon products *J. Mater. Chem.* A **10** 16171–81
[51] Gauthier J A, Stenlid J H, Abild-Pedersen F, Head-Gordon M and Bell A T 2021 The role of roughening to enhance selectivity to C_{2+} products during CO_2 electroreduction on copper *ACS Energy Lett.* **6** 3252–60
[52] Haastrup S *et al* 2018 The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals *2D Mater.* **5** 042002
[53] Gjerding M N *et al* 2021 Recent progress of the Computational 2D Materials Database (C2DB) *2D Mater.* **8** 044002
[54] Zhou J *et al* 2019 2DMatPedia, an open computational database of two-dimensional materials from top-down and bottom-up approaches *Sci. Data* **6** 86
[55] Satsangi S, Mishra A and Singh A K 2022 Feature blending: an approach toward generalized machine learning models for property prediction *ACS Phys. Chem.* **2** 16–22

[56] Hashimoto W, Tsuji Y and Yoshizawa K 2020 Optimization of work function via Bayesian machine learning combined with first-principles calculation *J. Phys. Chem.* C **124** 9958–70
[57] Rajan A C, Mishra A, Satsangi S, Vaish R, Mizuseki H, Lee K-R and Singh A K 2018 Machine-learning-assisted accurate band gap predictions of functionalized MXene *Chem. Mater.* **30** 4031–8
[58] Li N, Zong T and Zhang Z (eds) 2021 Prediction of the electronic work function by regression algorithm in machine learning *2021 IEEE 6th Int. Conf. on Big Data Analytics (ICBDA)* (*5–8 March 2021*)(https://doi.org/10.1109/ICBDA51983.2021.9403202)
[59] Kumar R and Singh A K 2021 Chemical hardness-driven interpretable machine learning approach for rapid search of photocatalysts *npj Comput. Mater.* **7** 197
[60] Wexler R B, Martirez J M P and Rappe A M 2018 Chemical pressure-driven enhancement of the hydrogen evolving activity of Ni_2P from nonmetal surface doping interpreted via machine learning *J. Am. Chem. Soc.* **140** 4678–83
[61] Lamoureux P S, Choksi T S, Streibel V and Abild-Pedersen F 2021 Combining artificial intelligence and physics-based modeling to directly assess atomic site stabilities: from sub-nanometer clusters to extended surfaces *Phys. Chem. Chem. Phys.* **23** 22022–34
[62] Waber J T and Cromer D T 1965 Orbital radii of atoms and ions *J. Chem. Phys.* **42** 4116–23
[63] Haynes W M (ed) 1978 *CRC Handbook of Chemistry and Physics* (Boca Raton, FL: CRC Press) (https://doi.org/10.1201/b17118)
[64] Linstrom P J and Mallard W G 2023 *NIST Chemistry WebBook (NIST Standard Reference Database Number 69)* (Gaithersburg, MD: National Institute of Standards and Technology) (https://doi.org/10.18434/T4D303) (available at: https://webbook.nist.gov/cgi/cbook.cgi?Contrib=)
[65] Pedregosa F *et al* 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30 (available at: http://jmlr.org/papers/v12/pedregosa11a.html)
[66] Chollet F 2021 Keras (available at: https://github.com/fchollet/keras)
[67] Stephens T 2016 Introduction to GP (available at: https://gplearn.readthedocs.io/en/latest/intro.html)
[68] Thygesen K S 2018 Computational 2D Materials Database (C2DB) (available at: https://cmr.fysik.dtu.dk/c2db/c2db.html)
[69] Le T A, Bui Q V, Tran N Q, Cho Y, Hong Y, Kawazoe Y and Lee H 2019 Synergistic effects of nitrogen doping on MXene for enhancement of hydrogen evolution reaction *ACS Sustain. Chem. Eng.* **7** 16879–88
[70] Giannozzi P *et al* 2009 QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials *J. Phys.: Condens. Matter* **21** 395502
[71] Hjorth Larsen A *et al* 2017 The atomic simulation environment—a Python library for working with atoms *J. Phys.: Condens. Matter* **29** 273002
[72] Perdew J P, Burke K and Ernzerhof M 1996 Generalized gradient approximation made simple *Phys. Rev. Lett.* **77** 3865–8
[73] Vanderbilt D 1990 Soft self-consistent pseudopotentials in a generalized eigenvalue formalism *Phys. Rev.* B **41** 7892–5
[74] Bengtsson L 1999 Dipole correction for surface supercell calculations *Phys. Rev.* B **59** 12301–4
[75] Monkhorst H J and Pack J D 1976 Special points for Brillouin-zone integrations *Phys. Rev.* B **13** 5188–92
[76] Momma K and Izumi F 2011 VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data *J. Appl. Crystallogr.* **44** 1272–6
[77] Alhabeb M, Maleski K, Anasori B, Lelyukh P, Clark L, Sin S and Gogotsi Y 2017 Guidelines for synthesis and processing of two-dimensional titanium carbide ($Ti_3C_2T_x$ MXene) *Chem. Mater.* **29** 7633–44
[78] Choksi T S, Roling L T, Streibel V and Abild-Pedersen F 2019 Predicting adsorption properties of catalytic descriptors on bimetallic nanoalloys with site-specific precision *J. Phys. Chem. Lett.* **10** 1852–9
[79] Dean J, Taylor M G and Mpourmpakis G 2019 Unfolding adsorption on metal nanoparticles: connecting stability with catalysis *Sci. Adv.* **5** eaax5101
[80] Mamun O, Winther K T, Boes J R and Bligaard T 2020 A Bayesian framework for adsorption energy prediction on bimetallic alloy catalysts *npj Comput. Mater.* **6** 177
[81] Wellendorff J, Silbaugh T L, Garcia-Pintos D, Nørskov J K, Bligaard T, Studt F and Campbell C T 2015 A benchmark database for adsorption bond energies to transition metal surfaces and comparison to selected DFT functionals *Surf. Sci.* **640** 36–44
[82] Montemore M M, Nwaokorie C F and Kayode G O 2020 General screening of surface alloys for catalysis *Catal. Sci. Technol.* **10** 4467–76
[83] Streibel V, Choksi T S and Abild-Pedersen F 2020 Predicting metal–metal interactions. I. The influence of strain on nanoparticle and metal adlayer stabilities *J. Chem. Phys.* **152** 094701
[84] Agresti A *et al* 2019 Titanium-carbide MXenes for work function and interface engineering in perovskite solar cells *Nat. Mater.* **18** 1228–34
[85] Jennings P C, Lysgaard S, Hummelshøj J S, Vegge T and Bligaard T 2019 Genetic algorithms for computational materials discovery accelerated by machine learning *npj Comput. Mater.* **5** 46
[86] Li B and Rangarajan S 2022 A conceptual study of transfer learning with linear models for data-driven property prediction *Comput. Chem. Eng.* **157** 107599
[87] Ghanekar P G, Deshpande S and Greeley J 2022 Adsorbate chemical environment-based machine learning framework for heterogeneous catalysis *Nat. Commun.* **13** 5788