

# Machine Learning-Driven Discovery of Key Descriptors for CO<sub>2</sub> Activation over Two-Dimensional Transition Metal Carbides and Nitrides

B. Moses Abraham, Oriol Piqué, Mohd Aamir Khan, Francesc Viñes,\* Francesc Illas, and Jayant K. Singh\*



Cite This: *ACS Appl. Mater. Interfaces* 2023, 15, 30117–30126



Read Online

ACCESS |

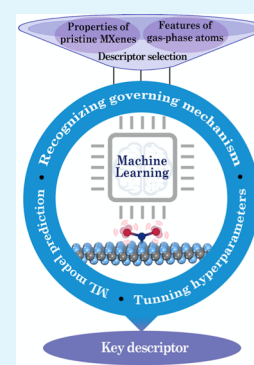
Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Fusing high-throughput quantum mechanical screening techniques with modern artificial intelligence strategies is among the most fundamental—yet revolutionary—science activities, capable of opening new horizons in catalyst discovery. Here, we apply this strategy to the process of finding appropriate key descriptors for CO<sub>2</sub> activation over two-dimensional transition metal (TM) carbides/nitrides (MXenes). Various machine learning (ML) models are developed to screen over 114 pure and defective MXenes, where the random forest regressor (RFR) ML scheme exhibits the best predictive performance for the CO<sub>2</sub> adsorption energy, with a mean absolute error  $\pm$  standard deviation of  $0.16 \pm 0.01$  and  $0.42 \pm 0.06$  eV for training and test data sets, respectively. Feature importance analysis revealed  $d$ -band center ( $\epsilon_d$ ), surface metal electronegativity ( $\chi_M$ ), and valence electron number of metal atoms ( $M_V$ ) as key descriptors for CO<sub>2</sub> activation. These findings furnish a fundamental basis for designing novel MXene-based catalysts through the prediction of potential indicators for CO<sub>2</sub> activation and their posterior usage.

**KEYWORDS:** MXenes, CO<sub>2</sub> activation, machine learning, density functional calculations, descriptors



## 1. INTRODUCTION

The excessive carbon dioxide (CO<sub>2</sub>) concentration in Earth's atmosphere has become a large threat to the environment given its main role in global warming; therefore, a lot of efforts have been taken worldwide to remove it. The rise of CO<sub>2</sub> concentration in the atmosphere is mainly due to the massive destruction of forests as well as the extensive exploitation of fossil fuels, which led to a continuous increase of CO<sub>2</sub> concentration, that will reach  $\sim 590$  ppm by the year 2100, causing an expected global temperature raise by 1.9 °C, with the concomitant acidification of oceans and devastating consequences for the marine ecosystems.<sup>1</sup> At present, the increasing CO<sub>2</sub> emissions are partly controlled through either converting it into useful carbon-based fuels/chemicals or by storing it in a stabilized media. To attain a valid impact on both environment and economy, it is necessary to utilize CO<sub>2</sub> instead of just storing it, to thus unlock its potential and trigger profitable industrial applications. Hitherto, several types of catalysts were investigated aimed at CO<sub>2</sub> activation and reduction, including different metal oxides, pure metals and alloys, organometallics, single-atom catalysts, non-metals, and nano-metals.<sup>2–4</sup> Typically, metals such as Cd, Sn, In, Pd, and Bi mediate the formation of formic acid from CO<sub>2</sub>,<sup>5–8</sup> while Ti, Zn, and Au can efficiently convert CO<sub>2</sub> into CO.<sup>9–11</sup> These studies demonstrated that the CO<sub>2</sub> molecule can interact with metal surfaces through either strong or weak binding modes. In the case of strong interactions, the metal–carbon (M–C)

overbonding may poison the catalyst surface, making the active sites inaccessible for further reduction of CO<sub>2</sub> with a concomitant reduction of product formation. In contrast, a weak bonding between CO<sub>2</sub> and a given metal surface does not allow for the CO<sub>2</sub> C–O bond dissociation, as desorption prevails over this bond scission chemical step, which would favor the formation of the desired products. It should be borne in mind that the C–O bond enthalpy in the CO<sub>2</sub> molecule is very large, of 803 kJ·mol<sup>−1</sup>,<sup>12</sup> and thus the activation of CO<sub>2</sub> can be regarded as a suitable approach to lower the CO<sub>2</sub> reaction conditions and energy demands. Therefore, a thorough activation analysis is highly required when designing novel catalysts based on rational approaches, to uncover which factors govern both activity and selectivity during the reactive processes.

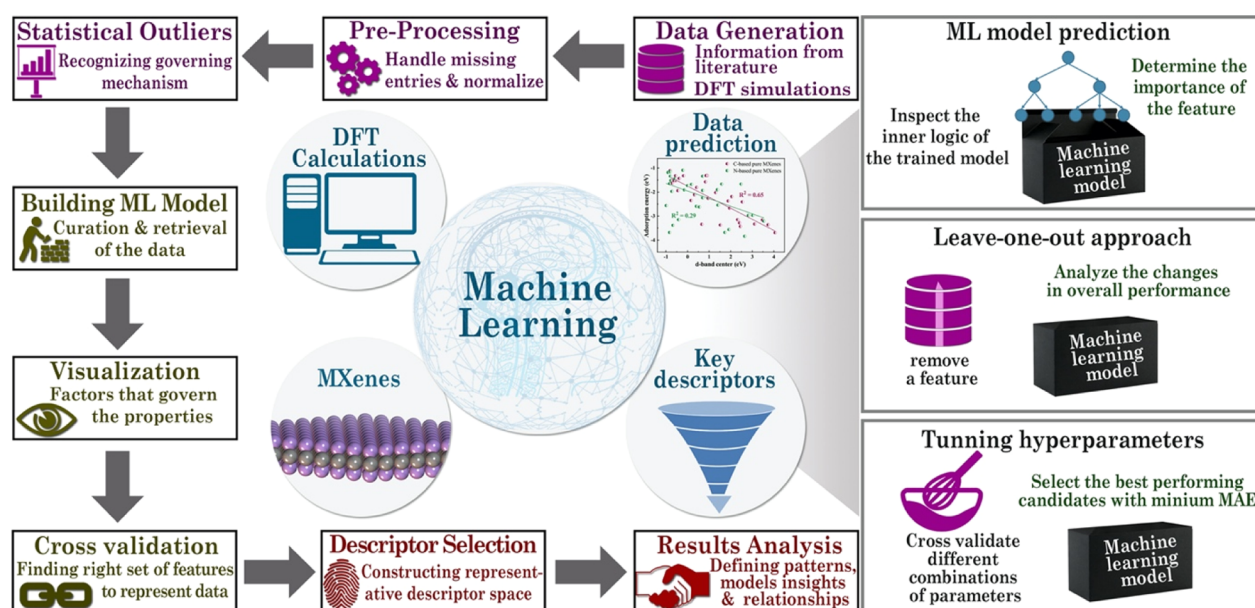
In general, the CO<sub>2</sub> binding energy over a potential catalytic surface is considered as an effective source for predicting the likelihood of CO<sub>2</sub> reduction reactions.<sup>13</sup> However, the experimental accurate measurement of the CO<sub>2</sub> binding energy is far from being a simple issue.<sup>14</sup> On the other hand,

**Received:** February 27, 2023

**Accepted:** May 16, 2023

**Published:** June 19, 2023





**Figure 1.** Schematic diagram of the ML workflow, trained on a data set generated from previous work and DFT calculations, aimed at identifying potential descriptors for CO<sub>2</sub> activation over two-dimensional TM carbides and nitrides (MXenes).

the theoretical modeling of the catalytic activity on a given material surface requires extensive yet accurate calculations, preferably from first principles-based methods, leading to a good understanding of the interaction of CO<sub>2</sub> with the surface of interest and accounting also for coverage effects, but at a high computational cost, though. In this regard, properties that provide information about the catalytic activity from a lower computational cost are highly preferable, particularly to screen over a pool of chemically related family of materials. In a simple case scenario, the adsorption energies can be linearly correlated with electronic descriptors that only require investigating the substrates, *e.g.*, by density functional theory (DFT), significantly reducing the computation cost for predicting the catalytic activity. In particular, Hammer and Nørskov<sup>15</sup> proposed one of the most successful descriptors to date, the *d*-band center, capable of predicting the adsorption energy of a given adsorbate at different TM surfaces using information of the TM surface electronic structure only. Here, the essence of the statement indicates that the binding energy of CO<sub>2</sub> to the TM surface does not require entire details of the density of states, where instead the *d*-band center is sufficient to correlate the interaction strength with the surface chemical activity and, eventually, the catalytic performance. In addition, other structural parameters such as bond lengths and angles, even surface coordination numbers, could be correlated with adsorption energies. By mapping the adsorption energy with materials intrinsic properties, one can obtain descriptors that do not only provide a fast screening over them with a rather high accuracy but also offer fundamental insights into the coupling between CO<sub>2</sub> and the surfaces of interest.

In the recent years, machine learning (ML) models trained on a limited number of quantum-mechanical calculations have become an appealing alternative for high-throughput prediction of chemical reactivity with either algorithm-derived<sup>16–20</sup> or handcrafted features.<sup>21–23</sup> The input variables required for ML modeling are typically accessible from the relaxed pristine materials surface structures without the presence of adsorbates. Using such properties with a low computational cost, one can predict complex parameters such

as catalytic activities or adsorption energy distributions in a much faster way. Since the ML analysis within the catalysis field mainly deals with particular chemical or physical properties, *e.g.*, adsorption energies, *d*-band centers, selectivities, limiting potentials, and so on, it is essential to consider supervised ML algorithms that map the target data set. Linear regression is a simple procedure with a highly potential and widespread approach used to analyze descriptors and to establish scaling relations for predicting valuable information in the computational heterogeneous catalysis field. More advanced techniques are currently available to handle multiple features such as non-linear relationships,<sup>24–27</sup> including kernel ridge regression,<sup>28</sup> neural networks,<sup>29</sup> random forest regression,<sup>30</sup> and Gaussian processes regression,<sup>31</sup> to name a few. Ultimately, choosing suitable descriptors is essential in any ML to regulate the prediction power and the learning efficiency.<sup>32</sup>

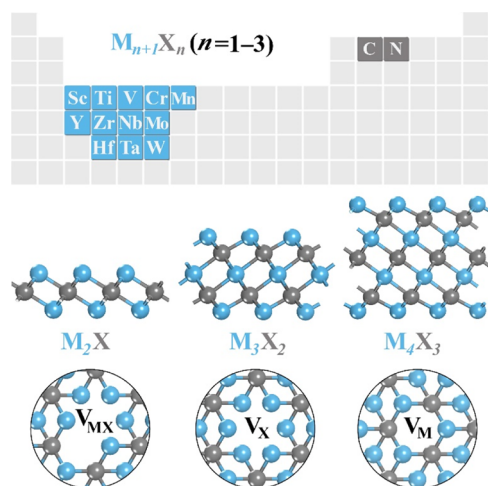
In this work, ML models are developed to mine and map the CO<sub>2</sub> activation over pure and defective MXenes based solely on their pristine properties and the features of gas-phase atoms that enter in the MXene chemical composition. Note in passing by that for CO<sub>2</sub> activation, we refer here to a strong interaction between the CO<sub>2</sub> molecule and the MXene surfaces, leading to significant changes in the adsorbed CO<sub>2</sub> geometry, including a bent geometry with elongated C–O bonds and a molecular negative charge, resulting from a charge transfer from the MXene surface to CO<sub>2</sub>. This CO<sub>2</sub> activation must not be mislabeled with another widely used meaning, implying the CO<sub>2</sub> conversion into other chemicals, *e.g.*, CO, formic acid, methanol, and so on, although both definitions are connected, since the bent CO<sub>2</sub> geometry is quite often the key, decisive state in CO<sub>2</sub> conversion, as found in organometallics,<sup>33</sup> TM carbides,<sup>34</sup> MXenes,<sup>35</sup> metals,<sup>36</sup> alloys,<sup>37</sup> and oxide-based catalysts.<sup>38,39</sup>

Thus, the fundamental goal of the present study is to develop and understand ML models for activated CO<sub>2</sub> adsorption on MXenes, which can be quantitatively implemented and leveraged for the predictive analysis in drawing useful information into the process of CO<sub>2</sub> posterior conversion. Figure 1 displays the schematic diagram of the

ML workflow, trained on a data set generated from our previous literature and DFT calculations to identify potential descriptors for CO<sub>2</sub> activation over MXenes. To this end, three regression models, namely, multivariate linear regression (MLR), decision tree regression (DTR), and random forest regression (RFR) are set up and evaluated with the aim of predicting potential descriptors for CO<sub>2</sub> activation over these materials.<sup>40</sup> Accordingly, we performed a feature importance evolution and investigated the effect of each primary feature on the target adsorption properties. As demonstrated below, the RFR model is best performing, using  $d$ -band center,  $\epsilon_d$ , the MXene surface metal electronegativity,  $\chi_M$ , and valence electron number of metal atoms,  $M_V$ , as meaningful features to predict the activation of CO<sub>2</sub> for the chosen MXene class of materials. This high-throughput screening research based on first-principles calculations and ML predictions can discover prominent indicators of CO<sub>2</sub> activation over MXene materials, and it is likely to be transferred to other bulk TM carbides/nitrides materials as well.

## 2. METHODOLOGY

**2.1. Data Collection and Pre-processing.** The data required to nurture the developed ML tools were collected from our previous literature on MXenes for CO<sub>2</sub> capture.<sup>35,41–44</sup> A total of 114 data points were extracted, among which 60 points are from pure MXenes with varying thickness, while remaining 54 points correspond to MXenes with different sorts of vacancies; see Figure 2. Note that



**Figure 2.** Side and top atomic structure views of TM carbides and nitrides MXenes with formula  $M_{n+1}X_n$  ( $n = 1-3$ ) with  $X = C$  or  $N$  and  $M$  metals from groups III–V of the Periodic Table. A total of 114 data points are extracted, among which 60 points are from pure MXenes with varying thickness, while remaining points correspond to MXenes with different vacancies; metal vacancy ( $V_M$ ), carbon/nitrogen vacancy ( $V_X$ ), and metal and nearby carbon/nitrogen vacancy ( $V_{MX}$ ).

MXenes are usually surface-terminated as a result of the synthesis procedure, yet bare MXenes are nowadays attainable either through molten salts synthesis<sup>45</sup> or after applying cleaning protocols.<sup>46</sup> Furthermore, such non-terminated sites have been appointed to be key catalytic active centers in CO<sub>2</sub> conversion, as shown in the dry methane reforming.<sup>47</sup> In addition, some previous cases, where \*CO<sub>2</sub> was found to dissociate into \*CO and \*O adsorbates upon relaxation on the MXene surfaces —due to a molecular placement too close to

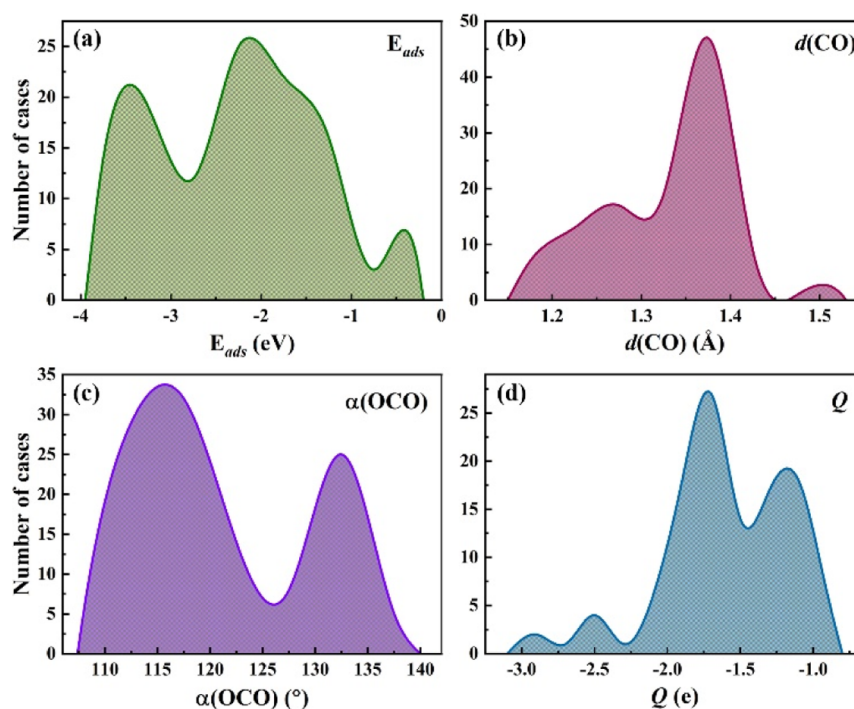
the MXene surface, and so a higher energy level, which led the dissociation— were reoptimized in order to gain a stable \*CO<sub>2</sub> adsorption state. In addition, we also observed that a substantial amount of data was missing, particularly on surface descriptors, which were here calculated and completed; see below. Notice that the data source had many aspects in common, *e.g.*, all being DFT calculations on  $p(3\times 3)$  slab models, with a minimum vacuum of 10 Å, and using Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional,<sup>48</sup> with Grimme’s D3 correction to account for dispersive forces.<sup>49</sup> However, data slightly differed concerning the plane-wave basis set kinetic energy-cutoff or the Brillouin zone  $k$ -points density. To assess the possible effect of such input differences on binding energies, we carried out test evaluations on 11% of the data set using the same materials, with representatives from pure MXenes and varying thicknesses and cases including different sorts of vacancies. The evaluated impact on target properties such as adsorption energies, bond lengths, and O=C=O angles were found to be at most of 0.07 eV, 0.03 Å, and 5.41°, respectively. Such discrepancies are well below or at least comparable to the inherent DFT accuracy.

The entire set of data points was then split into randomly selected training and test subsets. Accordingly, a random 20% of the total data points were labeled as test data and the remaining 80% was labeled as training data for the evaluation of the designed models. To better understand the importance of the studied models with the set of primary features, we considered the Pearson correlation coefficient,  $R$ , and the mean absolute error (MAE) as main evaluation indices.

**2.2. ML Models and Hyperparameter Tuning.** Three ML models, namely, MLR, DTR, and RFR, were devised and evaluated to predict the CO<sub>2</sub> activation over MXenes on the set of described input features or descriptors. Based on a training data set, each model was developed, where the test data set was employed to evaluate their prediction accuracy. For more detailed information about the considered three ML models; see Section S1 of the Supporting Information. To improve the model prediction quality, a cross-validation was carried out during the training process to tune the hyperparameters. Generally, the hyperparameter tuning (HT) is used for obtaining optimal model performance by finding a set of hyperparameters, which are tuned during the model training process,<sup>50</sup> *e.g.*, the DTR and RFR branches and leaf nodes; see Figures S1 and S2 in the Supporting Information. In the present study, HT was carried out using a grid search method, which is reliable methodology, while tuning a lower set of primary features. All the data processing and ML technique implementation were performed using the open-source scikit-learn library.<sup>51</sup>

**2.3. DFT Calculations.** Complementary periodic DFT calculations were carried out using the Vienna *ab initio* simulation package (VASP) code,<sup>52</sup> using a plane wave basis set for the valence electron density with an optimal kinetic energy cutoff of 415 eV. For the scalar-relativistic treatment of the effect of core electrons on the valence density, projector augmented wave<sup>53</sup> pseudopotentials were used. A generalized gradient approximation exchange–correlation functional has been employed, in particular, that proposed by Perdew–Burke–Ernzerhof (PBE).<sup>48</sup> The geometry optimization was considered converged when forces acting on atoms were all below 0.01 eV·Å<sup>−1</sup>, while an electronic convergence criterion of 10<sup>−5</sup> eV was imposed. An optimal Monkhorst–Pack grid of  $k$ -points





**Figure 3.** Density distribution of (a)  $\text{CO}_2$   $E_{\text{ads}}$ , (b) C–O bond distance,  $d(\text{CO})$ , (c)  $\text{CO}_2$  angle,  $\alpha(\text{OCO})$ , and (d) Bader charges,  $Q$ , for the complete set of data of 114 MXene cases.

of  $5 \times 5 \times 1$  dimensions was used, overall guaranteeing adsorption energies to be converged below chemical accuracy of  $1 \text{ kcal} \cdot \text{mol}^{-1}$ , *ca.* 0.04 eV. Dispersive forces were accounted using Grimme's D3 method,<sup>49</sup> being PBE-D3 a suited level of calculation employed in previous studies.<sup>41–44</sup>

The adsorption energy,  $E_{\text{ads}}$ , of  $\text{CO}_2$  on various MXene surfaces was obtained from the following equation

$$E_{\text{ads}} = E_{\text{CO}_2/\text{MXene}} - E_{\text{MXene}} - E_{\text{CO}_2} + \Delta E_{\text{ZPE}} \quad (1)$$

where  $E_{\text{CO}_2/\text{MXene}}$ ,  $E_{\text{MXene}}$ , and  $E_{\text{CO}_2}$  are the total energies of  $\text{CO}_2$  adsorbed on the corresponding MXene surface, that of the relaxed pristine MXene, and that of the isolated  $\text{CO}_2$  molecule, respectively. For the  $\text{CO}_2$  molecule, it was placed within a symmetric box of  $10 \times 10 \times 10 \text{ Å}$  dimensions and optimized at the  $\Gamma$ -point.  $\Delta E_{\text{ZPE}}$  is the zero point energy (ZPE) difference in between the adsorbed  $\text{CO}_2$  and that of the gas phase within the harmonic approximation. For further details, we refer to literature.<sup>35</sup>

As far as descriptors are concerned, the work function,  $\phi$ , is defined as the amount of energy required to move an electron from the material Fermi level,  $E_{\text{F}}$ , and place it in the vacuum energy level,  $E_{\text{vac}}$ . Thus

$$\phi = E_{\text{vac}} - E_{\text{F}} \quad (2)$$

In the  $d$ -band center model, it is defined as the gravimetric center of the  $d$ -projected density of states of a surface TM atom, within the initial energy level up to the energy level corresponding to an hypothetical  $d^{10}$  electronic configuration of the TM; see further details in literature.<sup>54</sup> Aside, a Bader's atoms-in-molecules electronic density analysis is carried out to integrate it within regions whose charge is assigned to certain atoms.<sup>55</sup> Thus, a negative  $Q$  value implies a negative charge, and *vice versa*. Finally, the exfoliation energies,  $E_{\text{exf}}$ , are gained, computed as the energy necessary to remove the A element from MXene MAX phase precursors,<sup>44</sup> and obtained as follows

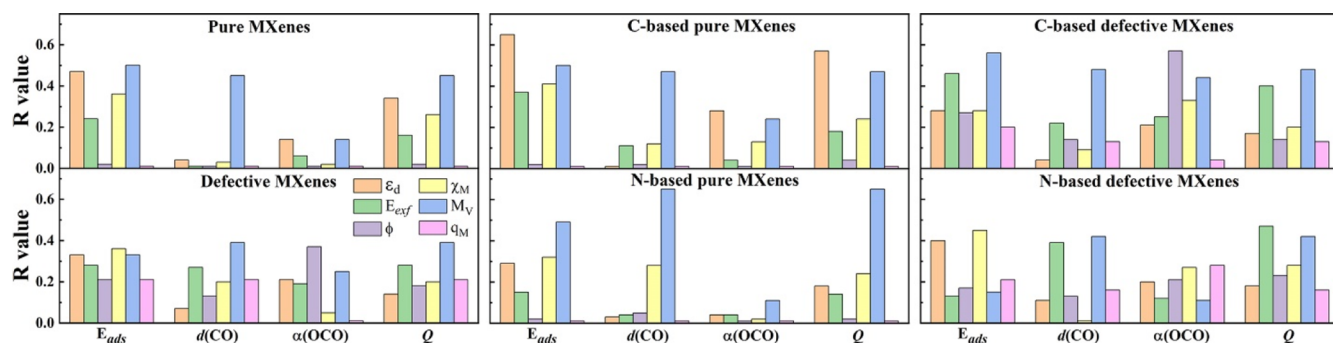
$$E_{\text{exf}} = (2 \cdot E_{\text{MXene}} + 2 \cdot E_{\text{A}} - E_{\text{MAX}}) / 4 \cdot S_{\text{A}} \quad (3)$$

where  $E_{\text{MXene}}$  and  $E_{\text{MAX}}$  are the isolated MXene and the MAX unit cell total energies, respectively, as depicted in Figure S3 of the Supporting Information. Besides,  $E_{\text{A}}$  and  $S_{\text{A}}$  indicate the bulk phase atomic energy of A species and the cross-section area of each created MXene unit, respectively. Within this definition, the larger the  $E_{\text{exf}}$ , the stronger the bonding between MXene layers and the A phase and the costlier is to separate them.

### 3. RESULTS AND DISCUSSION

Having consistently gained and gathered all the necessary data, we first considered four target variable indicators of the  $\text{CO}_2$  activation. These included  $\text{CO}_2$  adsorption energy,  $E_{\text{ads}}$ , in the sense that, *a priori*, the stronger the bonding, the higher the activation. Aside from this energetic feature, we regarded two geometric parameters, the average C–O bond distance,  $d(\text{CO})$ , and the  $\text{CO}_2$  molecular angle,  $\alpha(\text{OCO})$ , since, ideally, the activated  $\text{CO}_2$  features a reduced angle compared to the linear gas molecule angle of  $180^\circ$ , plus elongated C–O bonds, result from the activated bent geometry, and a consequence of a charge transfer from the substrate material.<sup>56,57</sup> Thus, the smaller the angle and the larger the bond lengths, the more activated the  $\text{CO}_2$ . Finally, the mentioned charge transfer can be quantified through the Bader charge of the adsorbed  $\text{CO}_2$ ,  $Q$ , in the sense that, the larger the charge, the more activated  $\text{CO}_2$  is.

At first, we evaluated these features in a descriptive fashion, showing fringe limits in the data set and distribution; see Figure 3. A quick inspection reveals that the distribution of features is not uniform for none of the target properties. For instance,  $E_{\text{ads}}$  shows three peaks, one close to *ca.*  $-3.5 \text{ eV}$ , another around  $-2.1 \text{ eV}$ , and a small peak close to  $-0.4 \text{ eV}$ . According to the Sabatier principle, moderate adsorption energies —neither too weak nor too strong— would provide



**Figure 4.** Regression coefficients,  $R$ , for the linear correlation between the target properties,  $E_{\text{ads}}$ ,  $d(\text{CO})$ ,  $\alpha(\text{OCO})$ , and  $Q$ , and primary features,  $\epsilon_d$ ,  $E_{\text{exf}}$ ,  $\phi$ ,  $\chi_M$ ,  $M_V$ , and  $q_M$ .

the better catalytic performances, but, in our case, an activated  $\text{CO}_2$  molecule getting bent and negatively charged often implies strong adsorption energies, suggesting that a surplus of energy is required for a reaction to occur when using adsorbed  $\text{CO}_2$ . In any case, among all the studied MXenes, only 6.84%—6 out of 114—of the  $E_{\text{ads}}$  are below  $-1.0$  eV, which indicates overall a strong  $\text{CO}_2$  chemisorption over the studied MXenes.

The previous property is accompanied by reduced angles and elongated C–O bonds, indicators of the  $\text{CO}_2$  activation.<sup>58</sup> In the latter case, they are concentrated at  $1.37$  Å, which is  $0.20$  Å larger than the  $\text{CO}_2$  distance in vacuum of  $1.17$  Å, with a smaller peak at  $1.27$  Å, and few cases with bond lengths larger than  $1.5$  Å, like  $\text{Cr}_2\text{C}$  with a  $d(\text{CO})$  of  $1.54$  Å. When it comes to molecular angles, there are two main peaks around  $116$  and  $132^\circ$ . Notice thus that all the studied cases imply a bent  $\text{CO}_2$ , with angles ranging from  $112$  to  $140^\circ$ . The increase in both C–O bond elongation and  $\text{CO}_2$  bending is consistent with a charge transfer from the surface to the adsorbed molecule.<sup>59</sup> Thus, the Bader charge of the adsorbed  $\text{CO}_2$  is also a potential indicator of activation, where the average  $Q$  is found to be  $-1.59$  e, with a minimum and maximum value of  $-2.98$  and  $-0.83$  e, respectively, and a significant peak around  $-1.1$  e.

To understand the efficiency of a catalyst, one requires descriptors that correlate with the catalyst performance. Hence, for a practical use, the selected primary features or descriptors should be much facile to evaluate when compared with that of the target properties and, whenever possible, connect with chemical intuition-derived concepts. Thus, for a fruitful comparison of unique fingerprints, we have considered 18 primary features aimed to characterize the local environment of the adsorption sites, chosen among the properties of pristine MXenes, but also including features from the atoms comprising the MXene. These primary features are rapid to obtain, unique, and easily accessible. Typically, since the binding energies scale linearly with the  $d$ -band filling, the adsorption strength could be linked to the TM  $d$ -band energy distribution. Figure S4 displays the linear correlations between the target properties;  $E_{\text{ads}}$ ,  $d(\text{CO})$ ,  $\alpha(\text{OCO})$ , and  $Q$ , and primary features of MXenes, including some of the best performing or alleged descriptors in the literature, such as the  $d$ -band center,  $\epsilon_d$ , the exfoliation energy,  $E_{\text{exf}}$ , the work function,  $\phi$ , the metal electronegativity,  $\chi_M$ , the valence electron number of a metal atom,  $M_V$ , and Bader charge of surface metal atom,  $q_M$ , along with the regression coefficients  $R$ . For a better understanding, the  $R$  values of the aforementioned descriptors are provided separately for pure

and defective MXenes and summarized in Figure 4, also regarding C- and N-based MXenes separately.

As seen in Figure 4, for both pure and defective MXenes,  $E_{\text{ads}}$  shows better linear trends with the primary features, while  $d(\text{CO})$  and  $\alpha(\text{OCO})$  show poor correlations when compared with other target properties. In the case of pure MXenes with varying thicknesses, the detailed analysis demonstrates that the  $R$  value of  $E_{\text{ads}}$  as a function of  $\epsilon_d$  improves by increasing it. For defective systems, the  $R$  value is smaller for single vacancies;  $V_M$  and/or  $V_X$ , while the  $R$  score increases in the case of double vacancies, i.e.,  $V_{\text{MX}}$ . Interestingly, the  $R$  value of several primary features exceeds that of the  $d$ -band center. For instance, in all cases,  $M_V$  shows better scaling relations among the other descriptors. It should be noted that the  $d$ -band center is quite a universal descriptor for  $E_{\text{ads}}$  of different adsorbates at transition metal surfaces representing catalyst models. However, there are several signatures that the  $d$ -band center itself is not an adequate descriptor for more complex compounds.<sup>60–63</sup> On the other hand,  $q_M$  shows very small regression coefficients, indicating that the target properties exhibit poor correlations with the primary features. Notably,  $M_V$  and  $\epsilon_d$  appear as the top two descriptors, independently establishing the relationship with the target properties. However, we were unable to establish a better relationship with target properties using simply several single descriptors, which requires integration of multiple descriptors to reach a more accurate description. Therefore, these insufficient correlations prompted us to build a predictive ML model through combination of primary features that could resemble the contribution of each feature individually to the model.

Thus, using primary features as input variables, we evaluated various ML models, including MLR, DTR, and RFR methods using our database. Their MAE together with the standard deviation,  $\sigma$ , are shown in Table 1. In the case of  $E_{\text{ads}}$ , the MAE values are found to be  $0.49 \pm 0.06$ ,  $0.53 \pm 0.10$ , and  $0.45 \pm 0.06$  eV for MLR, DTR, and RFR, respectively. Notice that

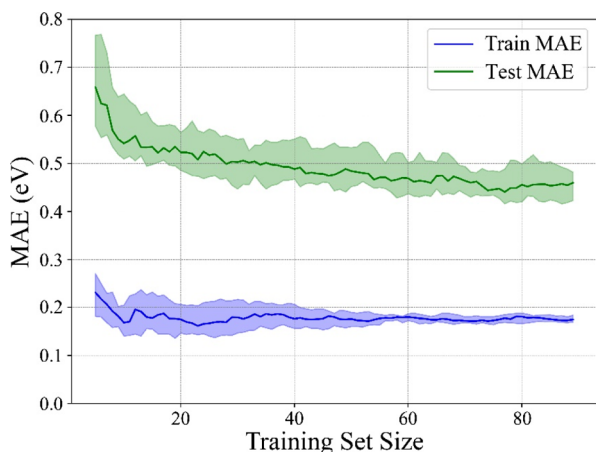
**Table 1.** MAE  $\pm$  Standard Deviation,  $\sigma$ , of  $E_{\text{ads}}$ ,  $d(\text{CO})$ ,  $\alpha(\text{OCO})$ , and  $Q$  Using MLR, DTR, and RFR ML Regressors, as Well as RFR RUF, and HT over RUF

ML model	$E_{\text{ads}}/\text{eV}$	$d(\text{CO})/\text{\AA}$	$\alpha(\text{OCO})/\text{deg}$	$Q/\text{e}$
MLR	$0.49 \pm 0.06$	$0.05 \pm 0.01$	$5.4 \pm 0.7$	$0.22 \pm 0.04$
DTR	$0.53 \pm 0.10$	$0.06 \pm 0.01$	$5.5 \pm 1.3$	$0.27 \pm 0.05$
RFR	$0.45 \pm 0.06$	$0.04 \pm 0.01$	$4.9 \pm 0.6$	$0.21 \pm 0.03$
RUF	$0.43 \pm 0.13$	$0.04 \pm 0.01$	$4.9 \pm 0.6$	$0.21 \pm 0.03$
HT	$0.42 \pm 0.06$	$0.04 \pm 0.01$	$4.8 \pm 0.8$	$0.20 \pm 0.03$

such errors are around double the typical DFT accuracy of *ca.* 0.2 eV and are still too large, especially when predicting cases with an  $E_{\text{ads}}$  weaker than  $-1$  eV. However, for the majority of MXene cases, the accuracy is already enough for a rapid screening, being the most of the cases between  $-1$  and  $-4$  eV; see Figure 1. For all the combinations of descriptors, the RFR model showed better performance than MLR and DFR models. Typically, feature importance estimates the weightage of a particular descriptor, thereby revealing the most relevant features for predicting the target properties by understanding the direct chemical insights. Especially for catalytic materials,<sup>30,64–67</sup> analyzing primary features is meaningful and interesting to predict the correlation between the target properties and the descriptors from view point of underlying physics and chemistry.

For  $\text{CO}_2$   $E_{\text{ads}}$ , the top five important features are the group number of the metal atom,  $G_{\text{M}}$ ,  $\chi_{\text{M}}$ ,  $M_{\text{V}}$ ,  $\epsilon_{\text{d}}$ , and  $E_{\text{exf}}$ . In the case of  $d(\text{CO})$ ,  $\alpha(\text{OCO})$ , and  $Q$ , the combinations of  $(\epsilon_{\text{d}}, E_{\text{exf}}, \phi, q_{\text{M}}$ , and  $\chi_{\text{M}})$ ,  $(\chi_{\text{M}}, \epsilon_{\text{d}}, q_{\text{M}}, \phi)$ , and number of d electrons,  $Nd_{\text{M}}$ , and  $(M_{\text{V}}, \chi_{\text{M}}, \phi, \epsilon_{\text{d}}$  and  $Nd_{\text{M}})$  were rendered as the top features. Among them,  $G_{\text{M}}$ ,  $\chi_{\text{M}}$ ,  $M_{\text{V}}$ , and  $Nd_{\text{M}}$  are tabulated chemical element properties, while  $\epsilon_{\text{d}}$ ,  $E_{\text{exf}}$ ,  $q_{\text{M}}$ , and  $\phi$  are DFT computed descriptors. To further understand the importance of precise features that correlate the target properties, it is necessary to remove the descriptors that are less relevant in minimizing the MAE. It should also be noted that an excessive number of features may lead to high prediction bias and low training efficiency.<sup>68</sup> To alleviate this issue, the feature dimension is reduced by employing the leave-one-out approach. Using this method, we eliminate unwanted features by evaluating their impact on the test set MAE. After shortlisting the descriptors according to the leave-one-out approach, HT was performed over RFR by employing cross-validation on various combinations of parameters. Although removing unnecessary features (RUF) and HT exhibited comparable performance, the latter marginally outperformed the former in terms of least MAE.

As per the size of the data set, Figure 5 shows the  $E_{\text{ads}}$  MAE decay with respect to training set size; in other words, the learning curve, regarding that training set considers randomly selected 80% of the samples, while the test set comprises the remaining 20%. For a better analysis, a cross-validation



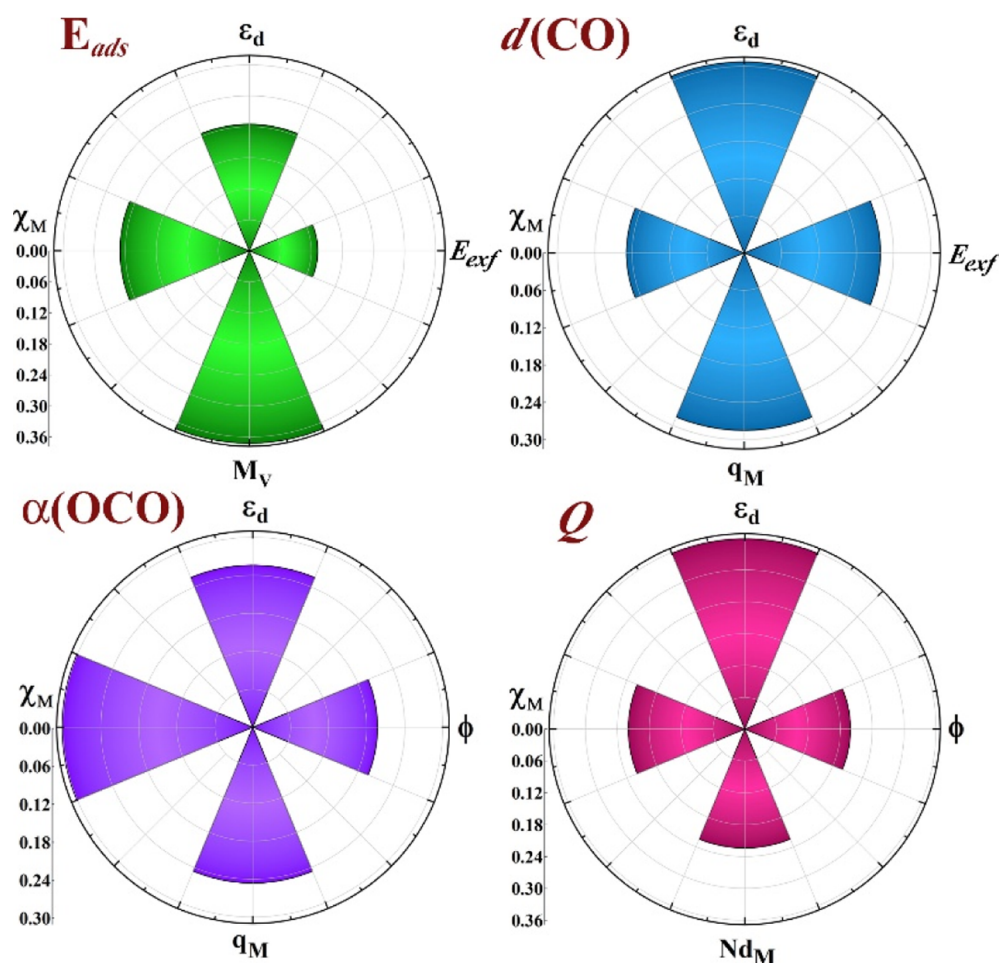
**Figure 5.** MAE evolution for the training (blue) and test set (green) versus training size for the prediction of  $E_{\text{ads}}$  using the HT of RFR ML algorithm. Shaded regions define the standard deviation limits.

procedure with 100 shuffle splits was carried out, as done in previous analysis, where average MAE is shown in Figure 5, with areas denoting the standard deviations.<sup>69</sup> Notice on the training set that RFR MAE decay is rather good,  $0.16 \pm 0.01$  eV, rapidly below the 0.2 eV DFT accuracy limit, and especially with an almost negligible standard deviation when having more than *ca.* 60 samples. Still, the decay of the test set is more pronounced, with larger standard deviations; see Table 1, and with the open question whether the evolution would remain stuck or would still descend when increasing the number of points of the data set. Alternatively, the reached plateau may be indicative of the existence of other descriptors, here not accounted for, which could be critical in improving the accuracy. Similar MAE evaluation is found for  $d(\text{CO})$ ,  $\alpha(\text{OCO})$ , and  $Q$  in Figures S5–S7 of the Supporting Information.

The HT of RFR further improved the accuracy of the model for  $E_{\text{ads}}$  by reducing the test set MAE to  $0.42 \pm 0.06$  eV. Indeed, estimations on the MAE on the HT of RFR ML using a limit training set of 113 points, and evaluated on the remaining test point, provides slightly better accuracies of 0.15 and 0.40 eV for training and test sets, respectively, over 114 developed ML models, signaling the convergence of the accuracy over the data set. The top four descriptors listed by the RFR model for  $E_{\text{ads}}$  are the combination of two features of the TM chemical elements,  $\chi_{\text{M}}$  and  $M_{\text{V}}$ , plus two other computed for the MXenes,  $\epsilon_{\text{d}}$  and  $E_{\text{exf}}$ ; see Figure 6, highlighting how important surface metal atoms are and how important is their placement within the MXene arrangement. It should also be noted that the choice of features introduces biasing, but at the same time, favors to counterbalance the overfitting, since we narrow their choice to sensible parameters that have been correlated to the sought, target properties, according to the literature. In the case of  $d(\text{CO})$  and  $\alpha(\text{OCO})$ , the MAE of the testing set is rather good as well, which are found to be  $0.04 \pm 0.01$  Å and  $4.84 \pm 0.78^\circ$ , respectively, essentially four times larger than chemical accuracy limits of 0.01 Å and  $1^\circ$ , respectively. For  $Q$ , there is a slight decrease in the prediction performance of RFR using HT; from  $0.21 \pm 0.03$  to  $0.20 \pm 0.03$  e, when compared to RUF. To reinforce the employed methodology, we have also compared our results using the recursive feature elimination (RFE)<sup>70</sup> method to filter the descriptors with extreme asymmetry (skewness) and with low/zero variance for recognizing more suitable smaller subset of features. As shown in Table S1 of the Supporting Information, the leave-one-out approach outperforms the RFE method by providing better predictive mean absolute errors. Finally, notice in Figure 6 that  $\epsilon_{\text{d}}$  and  $\chi_{\text{M}}$  are common descriptors of all the explored properties, while others such as  $\phi$ ,  $q_{\text{M}}$ , and  $E_{\text{exf}}$  are also common to a couple of properties, while  $M_{\text{V}}$  and  $Nd_{\text{M}}$  are only important to  $E_{\text{ads}}$  and  $Q$ , respectively.

Notice that the abovementioned ML models work irrespective of MXenes with or without vacancies, and for either C- or N-based MXenes, at variance with linear relationships; see Figure S4 of the Supporting Information, highlighting the versatility of the ML approach. Inspecting the descriptor weights in Figure 6, the ranking already states how  $\epsilon_{\text{d}}$  and  $\chi_{\text{M}}$  are determinant in  $\text{CO}_2$  activation, where the larger the  $\epsilon_{\text{d}}$ , the stronger the bonding is, as expected from the d-band model.<sup>19</sup> Aside, the smaller the metal electronegativity,  $\chi_{\text{M}}$ , the stronger the  $E_{\text{ads}}$ , fully physically understandable given the coulombic contribution of the bond of negatively charged





**Figure 6.** Feature importance of top four descriptors for  $E_{\text{ads}}$ ,  $Q$ ,  $d(\text{CO})$ , and  $\alpha(\text{OCO})$ .

$\text{CO}_2$  with positively charged surface metal atoms in the MXenes; see Figure S4 in the [Supporting Information](#). In any case, the weights of these two primary features are different for the different properties, e.g.,  $\epsilon_d$  weights are 30 and 35% for  $Q$  and  $d(\text{CO})$ , respectively, while for  $E_{\text{ads}}$ , actually  $\chi_M$  and  $\epsilon_d$  have similar importance values of 25%. Other secondary features can be rationalized as well; for instance, the  $\text{CO}_2$  charge  $Q$  also pretty much affects the molecular angle,  $\alpha(\text{OCO})$ , and is influenced by a smaller workfunction,  $\phi$ , which succinctly implies an easier MXene  $\rightarrow \text{CO}_2$  charge transfer.  $E_{\text{exf}}$  affects the bond strengths, and so, the larger the  $E_{\text{exf}}$ , the smaller the  $\text{CO}_2$  adsorption energy and the less elongated becomes  $d(\text{CO})$ . As far as geometries are concerned,  $d(\text{CO})$  and  $\alpha(\text{OCO})$  seem to be slightly influenced as well by  $q_M$ , so that the larger the charge, the smaller the  $\alpha(\text{OCO})$  angle and the longer the  $d(\text{CO})$ , stabilizing the negatively charged  $\text{CO}_2$ . Finally, the number of valence electrons and the number of d electrons, somehow related, affect the  $E_{\text{ads}}$  and the amount of transferred  $Q$ , in the sense that the smaller the number of valence electrons, and so, of d electrons, the stronger the  $E_{\text{ads}}$  and the more charge transferred, also in line with higher  $\epsilon_d$ . By identifying these descriptors, we have gained a deeper insight of the fundamental properties governing  $\text{CO}_2$  activation on the studied MXene surfaces, which can ultimately be used to design and optimize MXene-based compounds for  $\text{CO}_2$  storage or conversion applications. Thus, the ML tools allowed us to name which factors govern the  $\text{CO}_2$  activation, and which importance they have, which are properties to have in mind

when inspecting other MXenes for  $\text{CO}_2$  storage or usage selected processes. For example, from the descriptor weights in [Figure 6](#), when one would seek for  $\text{CO}_2$   $E_{\text{ads}}$  of  $-1$  eV or weaker, one should pay attention to the  $M_V$ ,  $\epsilon_d$ ,  $\chi_M$ , and  $E_{\text{exf}}$  descriptors; which is in line with the trends evaluated in [Figure S4](#) of the [Supporting Information](#); one would seek the MXene materials with  $\epsilon_d$  below  $-1$  eV, while having an  $E_{\text{exf}}$  above  $3.25 \text{ J} \cdot \text{m}^{-2}$ , a metal electronegativity of the metal,  $\chi_M$ , above 1.5, and a minimum number of 6  $e$  valence electrons of the metal,  $M_V$ . Moreover, the coefficient of determination analysis; see heat map in [Figure S8](#) of the [Supporting Information](#), demonstrates that the reduced set of features is sufficient for capturing the complex interactions influencing the  $E_{\text{ads}}$ ,  $d(\text{CO})$ ,  $\alpha(\text{OCO})$ , and  $Q$ , with no significant linear correlation among the found descriptors.

#### 4. CONCLUSIONS

In summary, we have developed a ML prediction scheme to unearth the potential indicators for  $\text{CO}_2$  activation on MXenes with the accessible properties of the pristine materials and of the atoms they are composed of. Three different ML algorithms were trained, where the hyperparameters tuning of RFR improved the accuracy of the model for  $E_{\text{ads}}$ , reducing the test set MAE to  $0.42 \pm 0.06$  eV when compared with that of the conventional RFR model, while the training set MAE was  $0.16 \pm 0.01$  eV. The high ranking of the  $d$ -band center,  $\epsilon_d$ , and surface metal electronegativity,  $\chi_M$ , is highlighted for  $E_{\text{ads}}$ , but also for other activation properties, including  $\text{CO}_2$  charge,

$Q$ ,  $d(\text{CO})$  bond length elongation, and molecular angle  $\alpha(\text{OCO})$  bending. These primary features are followed by valence and d electron numbers,  $M_V$  and  $Nd_M$ , and also MXene workfunctions,  $\phi$ , exfoliation energies,  $E_{\text{exf}}$  and surface metal charges,  $q_M$ , features, predicting the activation of  $\text{CO}_2$ , demonstrating the importance of such surface properties, and serving as a guide to select or search certain MXene materials for  $\text{CO}_2$  activation and/or use applications. Overall, the discovery of key descriptors for  $\text{CO}_2$  activation highlights the importance of ML strategies for accelerating the catalyst materials design and development by significantly extracting the information from a limited set of MXene database.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsami.3c02821>.

Description of the employed ML algorithms, including the MLR, the decision tree regressor, and the random forest regressor, graphical insights on how  $E_{\text{exf}}$  is calculated, lists of target properties and primary features, all linear correlations between target properties and primary features, distinguishing bare and defective MXenes, ML model learning curves, MAE when using the leave-one-out or the RFE procedures, coefficient of determination heat maps for the employed features, and references (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Francesc Viñes** – Departament de Ciència de Materials i Química Física, Institut de Química Teòrica i Computacional (IQTUB), Universitat de Barcelona, Barcelona 08028, Spain; [orcid.org/0000-0001-9987-8654](https://orcid.org/0000-0001-9987-8654); Email: [francesc.vines@ub.edu](mailto:francesc.vines@ub.edu)

**Jayant K. Singh** – Department of Chemical Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India; Prescience Insilico Private Limited, Bangalore 560049, India; [orcid.org/0000-0001-8056-2115](https://orcid.org/0000-0001-8056-2115); Email: [jayantks@iitk.ac.in](mailto:jayantks@iitk.ac.in)

### Authors

**B. Moses Abraham** – Department of Chemical Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India; Departament de Ciència de Materials i Química Física, Institut de Química Teòrica i Computacional (IQTUB), Universitat de Barcelona, Barcelona 08028, Spain; [orcid.org/0000-0002-1867-5239](https://orcid.org/0000-0002-1867-5239)

**Oriol Piqué** – Departament de Ciència de Materials i Química Física, Institut de Química Teòrica i Computacional (IQTUB), Universitat de Barcelona, Barcelona 08028, Spain; [orcid.org/0000-0002-6995-5927](https://orcid.org/0000-0002-6995-5927)

**Mohd Aamir Khan** – Department of Chemical Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India; Prescience Insilico Private Limited, Bangalore 560049, India

**Francesc Illas** – Departament de Ciència de Materials i Química Física, Institut de Química Teòrica i Computacional (IQTUB), Universitat de Barcelona, Barcelona 08028, Spain; [orcid.org/0000-0003-2104-6123](https://orcid.org/0000-0003-2104-6123)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acsami.3c02821>

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The work has been performed under the Project HPC-EUROPA3 (INFRAIA-2016-1-730897), with the support of the EC Research Innovation Action under the H2020 Program. J.K.S. and B.M.A. would like to thank the support of Science and Engineering Research Board (SERB), Government of India under the proposal MTR/2019/000194 and PDF/2021/000487, respectively. The research carried out at the *Universitat de Barcelona* has been supported by the Spanish MCIN/AEI/10.13039/501100011033 funded RTI2018-095460-B-I00, PID2021-126076NB-I00, and María de Maeztu MDM-2017-0767 grants, including funding from European Union and, in part, by and COST Action CA18234. We acknowledge Barcelona Supercomputing Center (BSC) and 'PARAM Sanganak' at IIT Kanpur for providing computing resources.

## ■ REFERENCES

- (1) Metz, B.; Davidson, O. R.; Bosch, P. R.; Dave, R.; Meyer, L. A. IPCC, 2007: Summary for Policymakers. In *Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate*; Cambridge University Press: Cambridge, UK and New York, NY, US, 2007.
- (2) Wang, Z. L.; Li, C.; Yamauchi, Y. Nanostructured Nonprecious Metal Catalysts for Electrochemical Reduction of Carbon Dioxide. *Nano Today* **2016**, *11*, 373–391.
- (3) Li, F.; Chen, L.; Xue, M.; Williams, T.; Zhang, Y.; MacFarlane, D. R.; Zhang, J. Towards a Better Sn: Efficient Electrocatalytic Reduction of  $\text{CO}_2$  to formate by Sn/SnS<sub>2</sub> derived from SnS<sub>2</sub> nanosheets. *Nano Energy* **2017**, *31*, 270–277.
- (4) Tian, Z.; Priest, C.; Chen, L. Recent Progress in the Theoretical Investigation of Electrocatalytic Reduction of  $\text{CO}_2$ . *Adv. Theory Simul.* **2018**, *1*, 1800004.
- (5) Gao, D.; Zhou, H.; Wang, J.; Miao, S.; Yang, F.; Wang, G.; Wang, J.; Bao, X. Size-Dependent Electrocatalytic Reduction of  $\text{CO}_2$  over Pd Nanoparticles. *J. Am. Chem. Soc.* **2015**, *137*, 4288–4291.
- (6) Chen, Z.; Wang, N.; Yao, S.; Liu, L. The Flaky Cd Film on Cu Plate Substrate: An Active and Efficient Electrode for Electrochemical Reduction of  $\text{CO}_2$  to Formate. *J. CO<sub>2</sub> Util.* **2017**, *22*, 191–196.
- (7) Luo, W.; Xie, W.; Li, M.; Zhang, J.; Zuttel, A. 3D Hierarchical Porous Indium Catalyst for Highly Efficient Electroreduction of  $\text{CO}_2$ . *J. Mater. Chem. A* **2019**, *7*, 4505–4515.
- (8) Bertin, E.; Garbarino, S.; Roy, C.; Kazemi, S.; Guay, D. Selective Electroreduction of  $\text{CO}_2$  to Formate on Bi and Oxide-Derived Bi Films. *J. CO<sub>2</sub> Util.* **2017**, *19*, 276–283.
- (9) Luo, W.; Zhang, J.; Li, M.; Zuttel, A. Boosting CO Production in Electrocatalytic  $\text{CO}_2$  Reduction on Highly Porous Zn Catalysts. *ACS Catal.* **2019**, *9*, 3783–3791.
- (10) Shi, C.; Anson, F. C. Catalysis of the Electro-Oxidation of Carbon Monoxide by Cobalt Octaethylporphyrin. *Inorg. Chem.* **2001**, *40*, 5829–5833.
- (11) Chen, Y.; Li, C. W.; Kanan, M. W. Aqueous  $\text{CO}_2$  Reduction at Very Low Overpotential on Oxide-Derived Au Nanoparticles. *J. Am. Chem. Soc.* **2012**, *134*, 19969–19972.
- (12) Glockler, G. Carbon–Oxygen Bond Energies and Bond Distances. *J. Phys. Chem.* **1958**, *62*, 1049–1054.
- (13) Abraham, B. M.; Parey, V.; Singh, J. K. A Strategic Review of MXenes as Emergent Building Blocks for Future Two-Dimensional Materials: Recent Progress and Perspectives. *J. Mater. Chem. C* **2022**, *10*, 4096–4123.
- (14) Hong, W. T.; Risch, M.; Stoerzinger, K. A.; Grimaud, A.; Suntivich, J.; Shao-Horn, Y. Toward the Rational Design of Non-Precious Transition Metal Oxides for Oxygen Electrocatalysis. *Energy Environ. Sci.* **2015**, *8*, 1404–1427.



- (15) Hammer, B.; Nørskov, J. K. Theoretical Surface Science and Catalysis—Calculations and Concepts. *Adv. Catal.* **2000**, *45*, 71–129.
- (16) Fung, V.; Hu, G.; Ganesh, P.; Sumpter, B. G. Machine Learned Features from Density of States for Accurate Adsorption Energy Prediction. *Nat. Commun.* **2021**, *12*, 88.
- (17) García-Muelas, R.; López, N. Statistical Learning Goes Beyond the *d*-Band Model Providing the Thermochemistry of Adsorbates on Transition Metals. *Nat. Commun.* **2019**, *10*, 4687.
- (18) Andersen, M.; Levchenko, S. V.; Scheffler, M.; Reuter, K. Beyond Scaling Relations for the Description of Catalytic Materials. *ACS Catal.* **2019**, *9*, 2752–2759.
- (19) Weng, B.; Song, Z.; Zhu, R.; Yan, Q.; Sun, Q.; Grice, C. G.; Yan, Y.; Yin, W.-J. Simple Descriptor Derived from Symbolic Regression Accelerating the Discovery of New Perovskite Catalysts. *Nat. Commun.* **2020**, *11*, 3513.
- (20) Back, S.; Tran, K.; Ulissi, Z. W. Toward a Design of Active Oxygen Evolution Catalysts: Insights from Automated Density Functional Theory Calculations and Machine Learning. *ACS Catal.* **2019**, *9*, 7651–7659.
- (21) Li, Z.; Wang, S.; Chin, W. S.; Achenie, L. E.; Xin, H. High-Throughput Screening of Bimetallic Catalysts Enabled by Machine Learning. *J. Mater. Chem. A* **2017**, *5*, 24131–24138.
- (22) Chowdhury, A. J.; Yang, W.; Walker, E.; Mamun, O.; Heyden, A.; Terejanu, G. A. Prediction of Adsorption Energies for Chemical Species on Metal Catalyst Surfaces Using Machine Learning. *J. Phys. Chem. C* **2018**, *122*, 28142–28150.
- (23) Tran, K.; Ulissi, Z. W. Active Learning across Intermetallics to Guide Discovery of Electrocatalysts for CO<sub>2</sub> Reduction and H<sub>2</sub> Evolution. *Nat. Catal.* **2018**, *1*, 696–703.
- (24) Goldsmith, B. R.; Esterhuizen, J.; Liu, J.-X.; Bartel, C. J.; Sutton, C. Machine Learning for Heterogeneous Catalyst Design and Discovery. *AIChE J.* **2018**, *64*, 2311–2323.
- (25) Schlexer Lamoureux, P.; Winther, K. T.; Torres, J. A. G.; Streibel, V.; Zhao, M.; Bajdich, M.; Abild-Pedersen, F.; Bligaard, T. Machine Learning for Computational Heterogeneous Catalysis. *Chemcatchem* **2019**, *11*, 3581–3601.
- (26) Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Deng, Z.; Ong, S. P. A Critical Review of Machine Learning of Energy Materials. *Adv. Energy Mater.* **2020**, *10*, 1903242.
- (27) Gu, G. H.; Choi, C.; Lee, Y.; Situmorang, A. B.; Noh, J.; Kim, Y.-H.; Jung, Y. Progress in Computational and Machine-Learning Methods for Heterogeneous Small-Molecule Activation. *Adv. Mater.* **2020**, *32*, 1907865.
- (28) Noh, J.; Back, S.; Kim, J.; Jung, Y. Active Learning with Non-Ab Initio Input Features toward Efficient CO<sub>2</sub> Reduction Catalysts. *Chem. Sci.* **2018**, *9*, 5152–5159.
- (29) Ma, X.; Li, Z.; Achenie, L. E. K.; Xin, H. Machine-Learning-Augmented Chemisorption Model for CO<sub>2</sub> Electroreduction Catalyst Screening. *J. Phys. Chem. Lett.* **2015**, *6*, 3528–3533.
- (30) Panapitiya, G.; Avendano-Franco, G.; Ren, P.; Wen, X.; Li, Y.; Lewis, J. P. Machine-Learning Prediction of CO Adsorption in Thiolated, Ag-Alloyed Au Nanoclusters. *J. Am. Chem. Soc.* **2018**, *140*, 17508–17514.
- (31) Li, Z.; Achenie, L. E. K.; Xin, H. An Adaptive Machine Learning Strategy for Accelerating Discovery of Perovskite Electrocatalysts. *ACS Catal.* **2020**, *10*, 4377–4384.
- (32) Abraham, B. M.; Sinha, P.; Halder, P.; Singh, J. K. Fusing a Machine Learning Strategy with Density Functional Theory to Hasten the Discovery of 2D MXene-based Catalysts for Hydrogen Generation. *J. Mater. Chem. A* **2023**, *11*, 8091–8100.
- (33) Jeoung, J.-H.; Dobbek, H. Carbon Dioxide Activation at the Ni<sub>4</sub>Fe-Cluster of Anaerobic Carbon Monoxide Dehydrogenase. *Science* **2007**, *318*, 1461–1464.
- (34) Sarabadani Tafreshi, S.; Panahi, S. F. K. S.; Taghizade, N.; Jamaati, M.; Ranjbar, M.; de Leeuw, N. H. Thermodynamic and Kinetic Study of Carbon Dioxide Hydrogenation on the Metal-Terminated Tantalum-Carbide (111) Surface: A DFT Calculation. *Catalysts* **2022**, *12*, 1275.
- (35) Morales-Salvador, R.; Gouveia, J. D.; Morales-García, Á.; Viñes, F.; Gomes, J. R. B.; Illas, F. Carbon Capture and Usage by MXenes. *ACS Catal.* **2021**, *11*, 11248–11255.
- (36) Peng, G.; Sibener, S. J.; Schatz, G. C.; Ceyer, S. T.; Mavrikakis, M. CO<sub>2</sub> Hydrogenation to Formic Acid on Ni(111). *J. Phys. Chem. C* **2012**, *116*, 3001–3006.
- (37) Behrendt, D.; Banerjee, S.; Clark, C.; Rappe, A. M. High-Throughput Computational Screening of Bioinspired Dual-Atom Alloys for CO<sub>2</sub> Activation. *J. Am. Chem. Soc.* **2023**, *145*, 4730–4735.
- (38) Mishra, A. K.; Roldan, A.; de Leeuw, N. H. CuO Surfaces and CO<sub>2</sub> Activation: A Dispersion-Corrected DFT+U Study. *J. Phys. Chem. C* **2016**, *120*, 2198–2214.
- (39) Nolan, M.; Fronzi, M. Activation of CO<sub>2</sub> at Chromia-Nanocluster-Modified Rutile and Anatase TiO<sub>2</sub>. *Catal. Today* **2019**, *326*, 68–74.
- (40) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (41) Parey, V.; Abraham, B. M.; Mir, S. H.; Singh, J. K. High-Throughput Screening of Atomic Defects in MXenes for CO<sub>2</sub> Capture, Activation, and Dissociation. *ACS Appl. Mater. Interfaces* **2021**, *13*, 35585–35594.
- (42) Jurado, A.; Ibarra, K.; Morales-García, Á.; Viñes, F.; Illas, F. Adsorption and Activation of CO<sub>2</sub> on Nitride MXenes: Composition, Temperature, and Pressure effects. *ChemPhysChem* **2021**, *22*, 2456–2463.
- (43) Morales-García, Á.; Calle-Vallejo, F.; Illas, F. Mxenes: New Horizons in Catalysis. *ACS Catal.* **2020**, *10*, 13487–13503.
- (44) Dolz, D.; Morales-García, Á.; Viñes, F.; Illas, F. Exfoliation Energy as a Descriptor of Mxenes Synthesizability and Surface Chemical Activity. *Nanomaterials* **2021**, *11*, 127.
- (45) Kamysbayev, V.; Filatov, A. S.; Hu, H.; Rui, X.; Lagunas, F.; Wang, D.; Klie, R. F.; Talapin, D. V. Covalent Surface Modifications and Superconductivity of Two-Dimensional Metal Carbide MXenes. *Science* **2020**, *369*, 979–983.
- (46) Persson, L.; Halim, J.; Lind, H.; Hansen, T. W.; Wagner, J. B.; Näslund, L. Å.; Darakchieva, V.; Palisaitis, J.; Rosen, J.; Persson, P. O. Å. 2D Transition Metal Carbides (MXenes) for Carbon Capture. *Adv. Mater.* **2019**, *31*, 1805472.
- (47) Kurllov, A.; Deeva, E. B.; Abdala, P. M.; Lebedev, D.; Tsoukalou, A.; Comas-Vives, A.; Fedorov, A.; Müller, C. R. Exploiting two-dimensional morphology of molybdenum oxycarbide to enable efficient catalytic dry reforming of methane. *Nat. Commun.* **2020**, *11*, 4920.
- (48) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (49) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (50) Piqué, O.; Koleva, I. Z.; Bruix, A.; Viñes, F.; Aleksandrov, H. A.; Vayssilov, G. N.; Illas, F. Charting the Atomic C Interaction with Transition Metal Surfaces. *ACS Catal.* **2022**, *12*, 9256–9269.
- (51) Scikit-learn, Machine Learning in Python. <https://scikit-learn.org/stable/index.html> (accessed March 24, 2022).
- (52) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations using a Plane-Wave Basis Set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169–11186.
- (53) Blochl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953–17979.
- (54) Vega, L.; Martínez, B.; Viñes, F.; Illas, F. Robustness of Surface Activity Electronic Structure-Based Descriptors of Transition Metals. *Phys. Chem. Chem. Phys.* **2018**, *20*, 20548–20554.
- (55) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, UK, 1990.

- (56) Kunkel, C.; Viñes, F.; Illas, F. Transition Metal Carbides as Novel Materials for CO<sub>2</sub> Capture, Storage, and Activation. *Energy Environ. Sci.* **2016**, *9*, 141–144.
- (57) Parey, V.; Abraham, B. M.; Jyothirmai, M. V.; Singh, J. K. Mechanistic Insights for Electrochemical Reduction of CO<sub>2</sub> into Hydrocarbon Fuels over O-terminated MXenes. *Catal. Sci. Technol.* **2022**, *12*, 2223–2231.
- (58) Boley, M.; Goldsmith, B.; Ghiringhelli, L. M.; Vreeken, J. Identifying Consistent Statements about Numerical Data with Dispersion-Corrected Subgroup Discovery. *Data Min. Knowl. Discov.* **2017**, *31*, 1391–1418.
- (59) Mazheika, A.; Wang, Y.-G.; Valero, R.; Viñes, F.; Illas, F.; Ghiringhelli, L. M.; Levchenko, S. V.; Scheffler, M. Artificial-Intelligence-Driven Discovery of Catalyst Genes with Application to CO<sub>2</sub> Activation on Semiconductor Oxides. *Nat. Commun.* **2022**, *13*, 419.
- (60) Xin, H.; Vojvodic, A.; Voss, J.; Nørskov, J. K.; Abild-Pedersen, F. Effects of *d*-Band Shape on the Surface Reactivity of Transition-Metal Alloys. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 115114.
- (61) Jennings, P. C.; Lysgaard, S.; Hansen, H. A.; Vegge, T. Decoupling strain and ligand effects in ternary nanoparticles for improved ORR electrocatalysis. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24737–24745.
- (62) Vojvodic, A.; Nørskov, J. K.; Abild-Pedersen, F. Electronic Structure Effects in Transition Metal Surface Chemistry. *Top. Catal.* **2014**, *57*, 25–32.
- (63) Gajdo, M.; Eichler, A.; Hafner, J. CO Adsorption on Close-Packed Transition and Noble Metal Surfaces: Trends from Ab Initio Calculations. *J. Phys.: Condens. Matter* **2004**, *16*, 1141–1164.
- (64) Frey, N. C.; Wang, J.; Vega Bellido, G. I.; Anasori, B.; Gogotsi, Y.; Shenoy, V. B. Prediction of Synthesis of 2D Metal Carbides and Nitrides (MXenes) and Their Precursors with Positive and Unlabeled Machine Learning. *ACS Nano* **2019**, *13*, 3031–3041.
- (65) Li, Z.; Wang, S.; Chin, W. S.; Achenie, L. E.; Xin, H. High-Throughput Screening of Bimetallic Catalysts Enabled by Machine Learning. *J. Mater. Chem. A* **2017**, *5*, 24131–24138.
- (66) Zhou, Q.; Tang, P.; Liu, S.; Pan, J.; Yan, Q.; Zhang, S. C. Learning atoms for materials discovery. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E6411–E6417.
- (67) Rajan, A. C.; Mishra, A.; Satsangi, S.; Vaish, R.; Mizuseki, H.; Lee, K.-R.; Singh, A. K. Machine-Learning-Assisted Accurate Band Gap Predictions of Functionalized MXenes. *Chem. Mater.* **2018**, *30*, 4031–4038.
- (68) Morooka, E. V.; Jager, M. O. J.; Federici Canova, F.; Himanen, L.; Foster, A. S.; Foster, A. S. Machine Learning Hydrogen Adsorption on Nanoclusters through Structural Descriptors. *npj Comput. Mater. Sci.* **2018**, *4*, 37.
- (69) Vega, L.; Viñes, F.; Neyman, K. M. Unravelling Morphological and Topological Energy Contributions of Metal Nanoparticles. *Nanomaterials* **2021**, *12*, 17.
- (70) Janet, J. P.; Kulik, H. J. Resolving Transition Metal Chemical Space: Feature Selection for Machine Learning and Structure–Property Relationships. *J. Phys. Chem. A* **2017**, *121*, 8939–8954.