Article

Can Machine Learning Predict the Reaction Paths in Catalytic CO$_2$ Reduction on Small Cu/Ni Clusters?

Rafał Stottko, Elżbieta Dziadyk-Stopyra [ID] and Bartłomiej M. Szyja *[ID]

Faculty of Chemistry, Wrocław University of Science and Technology, Gdańska 7/9, 50-344 Wrocław, Poland; 252633@student.pwr.edu.pl (R.S.)
* Correspondence: b.m.szyja@pwr.edu.pl; Tel.: +48-71-320-6370

Abstract: In this paper, we explore the catalytic CO$_2$ reduction process on 13-atom bimetallic nanoclusters with icosahedron geometry. As copper and nickel atoms may be positioned in different locations and either separated into groups or uniformly distributed, the possible permutations lead to many unnecessary simulations. Thus, we have developed a machine learning model aimed at predicting the energy of a specific group of bimetallic (CuNi) clusters and their interactions with CO$_2$ reduction intermediates. The training data for the algorithm have been provided from DFT simulations and consist only of the coordinates and types of atoms, together with the related potential energy of the system. While the algorithm is not able to predict the exact energy of the given complex, it is able to select the candidates for further optimization with reasonably good certainty. We have also found that the stability of the complex depends on the type of central atom in the nanoparticle, despite it not directly interacting with the intermediates.

1. Introduction

One of the most urgent issues facing human civilization is greenhouse gas emissions and the rise in the planet's average temperature [1]. According to The Global Monitoring Laboratory of the National Oceanic and Atmospheric Administration, the average concentration of CO$_2$ in the atmosphere in 2021 was 415 ppm, with an annual increase of 2.5 ppm [2]. CO$_2$ is not the most potent greenhouse gas, but it is the one that is found in the atmosphere in the greatest quantity. This has led to the investigation of several scenarios aiming to limit the warming of the planet to 1.5 °C, relative to the preindustrial period, or the reduction in long-term average temperature [3].

The direct capture of CO$_2$ from the air (Carbon Dioxide Removal (CDR)) followed by its chemical conversion to fuels or other energy carriers is often considered to be the most important solution, as it is able to close the carbon cycle and effectively make the branches of industry that rely on the fossil fuels zero-emission [4]. In order to carry out such a conversion, two components are necessary: highly active catalysts to increase the reaction rate and renewable energy source(s) to overcome the thermodynamic stability of CO$_2$ [5]. Electrocatalytic processes are often considered the most promising solutions, as they are able to satisfy both these requirements [6].

The main characteristics of a catalyst are its interactions with reactants, which is especially evident for Cu in electrocatalytic CO$_2$ conversion [7]. While other metals either deactivate rapidly due to a strong interaction with CO formed in the process or preferentially catalyze the proton/electron recombination without involving CO$_2$ [8], copper catalyzes the reduction of carbon dioxide to methane, ethylene [9–11] or even formic acid [7].

The main factor that determines selectivity towards a particular product is the reaction potential: with minimal potential, CO$_2$ does not react, and the end product is hydrogen. As the reaction potential increases, carbon monoxide or HCOOH can be obtained. The

highest reaction potential leads to the formation of hydrocarbons—mostly CH_4 and C_2H_4. The possibility of increasing the efficiency of this process upon designing a better catalyst makes the search for catalysts with more attractive parameters especially important.

Various forms of copper catalyst have been investigated, such as electrodes [9,11], foil electrodes [12], and copper nanotubes [10]; however, it is nanostructuring that offers an improvement in both activity and selectivity [13,14]. In the area of zeolite catalysis, the properties of copper nanoclusters have been investigated experimentally and computationally by Petranovskii et al. [15], with the conclusion that nanoclusters' geometry plays an important role. Often, the most stable structure for Me_{13} clusters is icosahedral, although, of course, this is not the universal rule. Thirteen-atom clusters with different stable geometries have been reported, e.g., for Ru_{13}, Rh_{13}, Os_{13}, or Ir_{13} [16].

As the number of atoms in the cluster increases, the average binding energy of these structures—and consequently their stability—also increases [17]. However, their activity in catalytic reactions decreases at the same time [18]. It is therefore desirable to find clusters of optimal size that are sufficiently stable and active. In the case of Cu_n clusters, the most stable structures are those where n = 13, 19, 23 (results common to many works) but also where n = e.g., 26, 55, 73, 92, 101, 115, 147 [17,19–21]. However, Grigoryan et al. found that the clusters with the Mackay icosahedra structure (n = 13, 55, 147) showed the greatest potential in terms of stability [19].

Conversely, the properties of metallic clusters are going to change when alloying them with other metals. For instance, Fe–Ni clusters with the icosahedral structure are more stable than pure, monometallic clusters, especially when iron occupies the central position in the structure and nickel atoms are located on the surface of the nanoparticle [22]. In addition, a small amount of Pd added to the copper catalyst increases the selectivity of the partial acetylene hydrogenation reaction catalyzed by this system [23], or Pt_nAg_n clusters are the most stable when most of the Pd is in the core and most of the Ag is on the shell of the cluster [24]. In such situations, both cluster composition and metal distribution can play a pivotal role, which complicates the issue of finding optimal clusters. In most cases, however, this is an effort that is worthwhile due to the better properties of bimetallic structures compared with monometallic alternatives.

2. Aim and Scope

In this work, we present a protocol for predicting the energy of bimetallic nanoparticle complexes with reactants. Shim et al. [25] showed that Cu_{13} shows the highest catalytic activity, as it has the lowest activation barrier for the reduction of CO_2 to CH_4 among Cu_{13}, Cu_{55}, and the Cu_{111} surface. Thus, we decided to focus on the Cu_nNi_{13-n} nanocluster interacting with CO_2 and the intermediates and products of its conversion. We present a comparison of the efficiency of simple and fast machine learning training methods. Our aim is to reduce as much as possible the time required to perform the optimization of individual structures, and therefore, we focus primarily on the geometries of the nanoclusters. This is a fairly rare approach, and it has not been used frequently so far, with only a few exceptions [26–28]. Predicting the energy of complexes based on their physicochemical properties is used more often, as the prediction accuracy of such models is essentially higher [29–36]. This approach, however, requires either careful optimization of each individual structure to obtain input data or to retrieve such data from external, already prepared databases, which excludes determining the predicted parameter for structures never studied before.

On top of that, the vast majority of the work we have examined focuses on predicting the properties of organic molecules, having distinct functional groups [34,36–41]. Significantly fewer reports predicting the properties of systems containing transition metals are available, likely due to the much greater complexity needed to describe the nature of the interactions in such systems [26–30,33,42].

In this work, we propose a procedure that predicts the catalytic activity of the systems containing Cu and Ni, using only geometric properties as descriptors, together with simple

and fast learning models. Our main goal was to create a tool that at the very least correctly determines the trends between groups of systems in order to designate the most promising candidates for a more detailed and time-consuming analysis.

3. Methodology

3.1. Models of the Nanoclusters

Nanoclusters with varying Cu:Ni ratios (Cu_nNi_{13-n}) were selected for analysis. The clusters were put in a periodic box (periodic boundary conditions), with the dimensions $24 \times 24 \times 24 \times$ Å and the angles $\alpha = \beta = \gamma = 90°$.

For all nanocluster–intermediate complexes, several types of binding modes were investigated, keeping the same distribution of the Cu and Ni atoms within the cluster, resulting in a total of 254 configurations.

All calculations were carried out within the DFT (Density Functional Theory) framework, as implemented in the VASP (Vienna Ab-initio Simulation Package) version 5.4.4 [43,44]. The exchange–correlation energy was defined by making use of the Perdew–Burke–Ernzerhof functional [45,46]. The plane wave energy cutoff was established at 450 eV, and the electron–ion interactions were characterized by the projector-augmented wave method [47].

3.2. Electrocatalytic Process Simulation

The basis for examining the electrocatalytic process is the method which relates the transfer of a proton/electron pair to the potential of the hydrogen electrode at equilibrium in the standard conditions [7]. The free energies for each electroreduction step are described by the following formula:

$$\Delta G = \Delta G_{product} - \Delta G_{reactant} - eU \tag{1}$$

where ΔG is the Gibbs free energy, e is elementary charge, and U is the applied potential.

The procedure assumes a linear relation between the applied potential and free energies. The potential limiting stage of the reaction step was defined as the minimum potential necessary to make all the consecutive reaction stages preserve the decreasing energy trend.

Importantly, the analysis focuses exclusively on the thermodynamics of the intermediate conversion and their interactions with the cathode surface, leaving out phenomena that do not contribute to the calculation of the potential of the process—such as reactant adsorption or product desorption.

3.3. Machine Learning

Determining the energy of all clusters, with every statistically possible distribution of metals, would be an extremely time-consuming task. If we begin to take into account all the possible ways of distributing Cu and Ni atoms in the studied clusters, the ways of binding intermediates to them, or rotating these intermediates, even with the simplifying assumptions that we made, the number of combinations possible to study exceeds 20 thousand. Therefore, in this work, we abandoned the optimization of all possible structures and instead developed a procedure aiming at the prediction of the stability of a particular system. The Procedure to Indicate Clusters Energy (PICle) at the lowest possible computational cost identifies a subset of the most preferred systems.

The procedure is coded in Python 3 language, and all the scripts associated with it are available at https://github.com/all2gos/PICle (accessed on 25 November 2023).

We tested several machine learning models, included in the scikit-learn library [48]: Support Vector Regression Machine [49], Kernel Ridge Regression [50], and Ensemble Methods (Random Forest [51,52], AdaBoost [53], and Extra Trees Regressor [54]). We selected the most popular representatives of the kernel and ensemble methods in order to test the accuracy of a large number of models following the no-free-lunch theorem.

The hyperparameters tested for each method are gathered in Table 1.

Table 1. Machine learning methods with hyperparameters used in this work.

ML Method	Hyperparameters	Search Ranges
Support Vector Regressor (svr)	kernel type regularization parameter C	linear, polynomial, sigmoidal, rbf 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2, 3, 4, 5, 8, 12, 15, 20
Kernel Ridge Regression (krr)	kernel type alpha value C width of no penalty tube (epsilon)	linear, polynomial, sigmoidal, rbf 0.2, 0.4, 0.6, 0.8, 1,1. 5, 2, 3, 4, 5, 8, 12, 15, 20 from 0.1 to 1.2 in 0.1 increments
Random Forest * (rf)	max. N of estimators min. N of samples to split a node max. depth of the tree	from 25 to 500 in 25 increments 2, 3, 4, 5, 6 none, 2, 3, 4, 5, 6, 7, 8
Extra Trees Regressor * (etr)	max. N of estimators min. N of samples to split a node max. depth of the tree	as in Random Forest as in Random Forest as in Random Forest
AdaBoost * (ada)	max. N of estimators learning rate loss function	as in Random Forest 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2, 3, 4, 5, 8 linear, square, exponential

* Random Forrest, Extra Trees Regressor, and AdaBoost belong to Ensemble Methods.

Ideally, the algorithm we developed would be capable of predicting the energies of the investigated systems; however, achieving this might not be possible for a method relying only on the geometric factors. The next best thing is to capture the trends, and in order to achieve this, we use rank correlation, which we call EI, for *Entropy Index*.

EI is a result of a comparison of a set that contains two subsets of values: prediction and validation. The values are sorted by increasing order in the predicted and validation subsets. Ideally, we would expect that for the particular cluster, its position in the prediction and validation sets would be exactly the same. This method allows for obtaining information about the order of the particular systems in these subsets with respect to increasing energy. EI is the average difference between the position of a given system in each sorted subset. In addition to the average, we also consider the maximum value. Finally, we divide these values by the length of the given dataset to make it independent of its length and convenient for comparisons between sets of different lengths.

4. Energy Predictions from Machine Learning

4.1. Training Model

To be able to effectively create test and training sets, we generated all possible combinations of the Cu and Ni distribution within the cluster. In order to accomplish this task, a simplification was made by assuming that each 13-atom CuNi cluster is characterized by ideal icosahedral geometry. This allowed us to represent each cluster using a one-dimensional representation (details on the notation created for this purpose can be found in the Supplementary Information).

CO_2 reduction can take place in one of three ways (Figure 1), and a previous study by Peterson and coworkers [7] concluded that these are the pathways requiring the lowest overpotential on Cu surface. We investigated all three pathways, with the assumption that regardless of the cluster composition and binding mode of the reactants, it is possible for all three to occur. In other words, we did not exclude any pathway from the analysis.

$$CO_2 \longrightarrow HCOO \longrightarrow HCOOH$$
$$\searrow COOH \longrightarrow CO$$
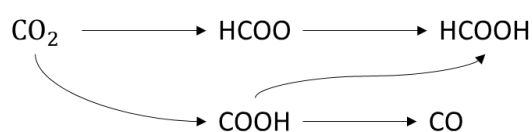
Figure 1. Possible conversion reaction paths CO_2.

In order to simplify an already complex analysis, we decided to limit it to only a few well-defined binding modes of the intermediates to the cluster. We did not take into account geometric distortions that are possible due to interactions with intermediates. We distinguished 5 different binding modes for the systems, which are shown in Figure 2. Note that the binding mode was not directly used as a descriptor; thus, it is possible to determine the energy of complexes with slightly different geometries without any modification of the procedure.

(**a**)	(**b**)	(**c**)	(**d**)	(**e**)
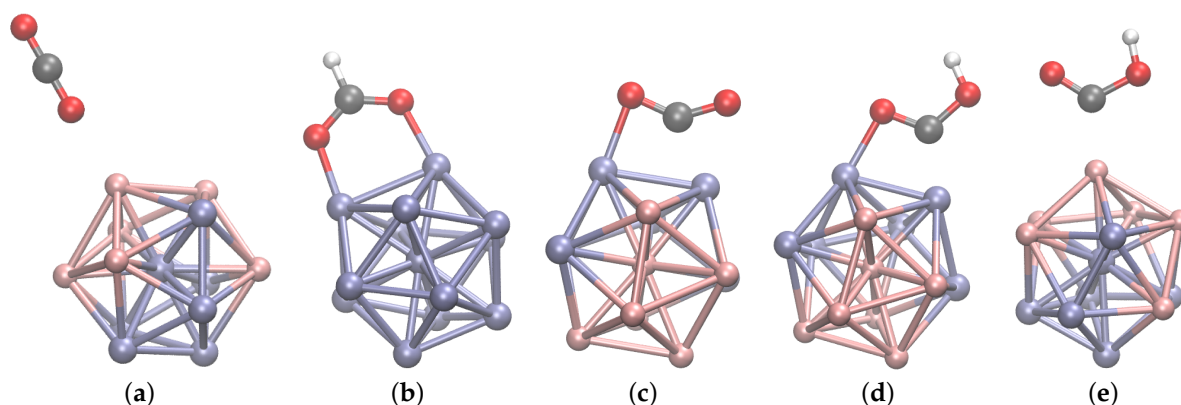
Figure 2. Binding modes of CO_2 and other intermediates to the cluster: (**a**) MO, (**b**) MOCOM, (**c**) MOCM, (**d**) MOCM_H_n_bond, (**e**) MC. Letters indicate dominant interaction(s); M stands for metal, either Cu or Ni.

4.2. Descriptors

There are several descriptors allowing each individual cluster to be represented numerically. As it was mentioned above, all descriptors presented in this work are based on the geometries of the clusters. Each cluster was described with the following characteristics:

1. The chemical composition of the cluster is represented by the Ni:Cu ratio, which is equivalent to the number of Ni (or Cu) atoms in the cluster. We used the number of Ni atoms in further work.
2. The distribution of metal atoms in the cluster can be represented either by a vector between the geometric center of the cluster and its center of mass or the number of Ni–Ni, Cu–Cu, and Cu–Ni bonds in the cluster.
3. The shortest paths, which is another way to describe the degree of dispersion of metal atoms in clusters. This refers to the number of atoms forming a directly connected chain, such as the path passing through the chain leading through each of its atoms exactly once.

    For the purposes of this project, 4 components were investigated:

    (a) The shortest path on outer Ni atoms, not including the central atom.
    (b) Th shortest path on outer Cu atoms, not including the central atom.
    (c) The shortest path along all Ni atoms, including the central atom if it is Ni, and its inclusion leads to a shorter path length.
    (d) The shortest path along all Cu atoms, including the central atom if it is Cu, and its inclusion leads to a shorter path length.

    These descriptors, however, exhibit a high correlation with each other, and training the model using all of them would not lead to any improvement in accuracy. Therefore, for further investigations, we decided to select the descriptor that was the most correlated with all the values we wanted to predict: the shortest path on outer Cu atoms.

4. The composition of the coordination sphere—the atom(s) of the cluster to which the intermediate is bound, consisting of several components: 0%Ni, 8%Ni, 16%Ni, 25%Ni, 33%Ni, 41%Ni, 50%Ni. Each of the described parameters takes the value $n \in <0, 13>$,

where $n \in \mathbf{N}$. For example, the value k for parameter 41%Ni means that in a given cluster, it is possible to distinguish k atoms of the metal, in which nickel atoms fill 41% of the first coordination sphere (5 atoms).

4.3. The Relevance of the Descriptors

Obviously, the energy of the system is mostly dependent on the chemical composition of the system. A system containing more Ni atoms in the cluster, or hydrogen atoms in the intermediate, will have lower potential energy. These descriptors are used by the machine learning model as the main determinant in predicting the energy of the system. Such a model would mostly capture the effect of the cluster composition as the main component of the final prediction. Other, more subtle effects, such as binding modes for instance, would be overshadowed, as their contribution to the total energy is smaller; however, they are of crucial importance in the catalytic effect. In other words, the most important differences that the model should capture are less dependent on the composition but rather indicate the relative energy difference between systems.

However, the composition of the system still carries important information, as the interactions of the intermediates depend on the composition of the cluster. Therefore, we could not completely disregard the properties of the system that are related to its composition, i.e., the number of nickel atoms and the presence of hydrogen in the system, but only reduce their importance in the overall prediction of energy. For this reason, we decided to test the following solutions:

1. In order to not take into account the effect of the different atomic composition of the cluster itself, a conversion of energy values was performed according to the following formula [22]:

$$E_{mix} = \frac{1}{13}\left(E_{Cu_n Ni_{13-n}} - \frac{1}{13}\left(nE_{Cu_{13}} + (13-n)E_{Ni_{13}}\right)\right) \tag{2}$$

    where:

    * $E_{Cu_n Ni_{13-n}}$, $E_{Cu_{13}}$, $E_{Ni_{13}}$ is the energy of clusters of corresponding atomic compositions expressed in eV.
    * n is the number of copper atoms in the system.

    The resulting parameter is called *mixing energy* and is denoted in the dataset as *mix_energy*.
2. The energy of the hydrogen molecule was determined under the same conditions under which the energy of all clusters was determined (Section 3.2). Then, half of this value was subtracted from the systems that contained a hydrogen atom in their composition. The energy thus determined is named in the dataset as *energy_wth_h2* and in the rest of the paper is abbreviated as *wth* or *energy_wth*. Note that it does not matter whether the energy_norm or energy_wth is used in Equation (2), as mathematically, they lead to the same result.

Thus, we identified three values for which we evaluate the predictive ability of the machine learning model: mixing energy (mix_energy), energy corrected for hydrogen present (energy_with_h2), and potential energy without any corrections (energy_norm).

4.4. Training and Test Sets

In order for the model to be trained, it is necessary to create datasets containing numerous cases on the basis of which training can be accomplished. Thus, we determined the energies of a significant number of complexes as accurately as possible with minimal computational effort.

The first step was to determine the energy of monometallic complexes with molecules bound in a specific mode. The DFT energies of these systems were determined as described in Section 3.2. In order to keep the computational time to a minimum, the energy of each different system that includes a bimetallic cluster was determined as a weighted average

of the energies of the monometallic clusters. This essentially assumes equal interaction between the metal atoms, regardless of metal type.

To check the validity of this assumption, the energy determined in this approximation was compared with the energy of the clusters that are part of the validation set in Figure 3). On the tested dataset, the EI of the system is only 0.012, with the maximum difference for a single case equal to 0.077. Therefore, it can be concluded that the simplifications we made are acceptable.
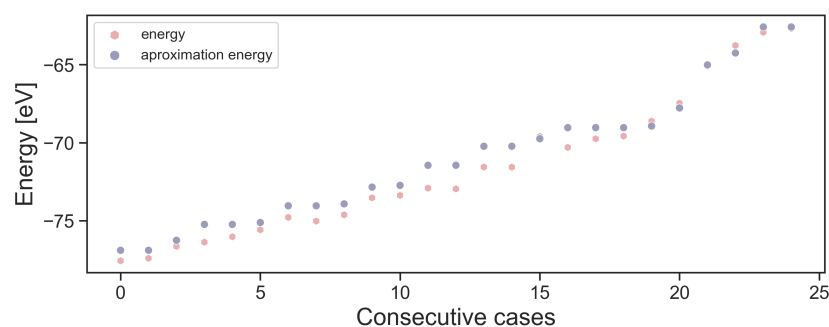
Figure 3. Comparison of energies determined by optimization of complexes with those determined by the proposed approximation techniques.

The set of energies of the complexes, determined using this technique, contained 649 elements. These values were used to train the machine learning model. We used a 70/30 split on the training/testing set for all tested models. To maximize the repeatability of the results, we set the random state to 42 [55]. We tested different training settings: firstly, we used a dataset that contained systems of any composition (all possible numbers of Ni atoms—*allni*); secondly, we restricted the model to clusters in which nickel did not account for more than 50% of the atoms (*ni7*). In addition, we checked the accuracy of the models depending on whether outliers were deleted (*2z*) or not (*0z*), with the outlier defined as a case whose Z-score in at least one descriptor is more than 2. In addition, as mentioned in Section 4.3, we tested the ability of the models to predict 3 values: energy_norm, energy_wth, and mix_energy. The combination of all possible settings of these parameters results in 12 training settings, which are gathered in Table 2, together with the calculated EI values.

5. Results and Discussion

5.1. Direct Value Prediction

Initially, we attempted to predict energy values directly. The EI values of the best model for each training method are gathered in Table 2, and the hyperparameters of the respective models are shown in Table 3.

The resulting accuracy is not very high, as can be seen from the relatively high EI values. The best value was observed for total energy corrected for hydrogen, trained for the set including all Ni:Cu ratios, not disregarding the outliers. The value is still relatively high and amounts to 0.079, which means the position in the validation set and set with predicted values differ on average by 7.9% of the total size of the set. The maximum EI values are in the range of approximately 0.25 to 0.50, which means the position differs by almost half of the length of the set.

Table 2. The accuracy of prediction depending on different training settings.

Model	Predicted Value	Conditions	Outliers	EI (max. EI) Test Set	EI (max. EI) Validation Set	MAE Test Set	MAE Validation Set
allni_0z_mix	mix_energy	Ni = all	no	0.089 (0.482)	0.188 (0.588)	0.015	0.024
allni_2z_mix	mix_energy	Ni = all	Z–score > 2	0.081 (0.390)	0.202 (0.632)	0.023	0.017
ni7_0z_mix	mix_energy	Ni < 7	no	0.094 (0.376)	0.170 (0.609)	0.012	0.021
ni7_2z_mix	mix_energy	Ni < 7	Z–score > 2	0.106 (0.483)	0.152 (0.551)	0.016	0.015
allni_0z_wth	energy_wth_h2	Ni = all	no	0.079 (0.246)	0.025 (0.074)	0.593	0.933
allni_2z_wth	energy_wth_h2	Ni = all	Z–score > 2	0.128 (0.435)	0.157 (0.441)	0.681	3.788
ni7_0z_wth	energy_wth_h2	Ni < 7	no	0.104 (0.398)	0.152 (0.609)	0.797	2.969
ni7_2z_wth	energy_wth_h2	Ni < 7	Z–score > 2	0.106 (0.333)	0.113 (0.412)	0.924	3.942
allni_0z_norm	energy	Ni = all	no	0.086 (0.431)	0.108 (0.348)	0.774	0.990
allni_2z_norm	energy	Ni = all	Z–score > 2	0.102 (0.435)	0.179 (0.536)	0.403	2.264
ni7_0z_norm	energy	Ni < 7	no	0.086 (0.459)	0.088 (0.217)	0.523	8.926
ni7_2z_norm	energy	Ni < 7	Z–score > 2	0.106 (0.529)	0.136 (0.406)	1.464	3.864

This observation is also visible in the predicted and calculated energy plots (see Figure 4). The best correspondence can be seen for the energy_norm predicted using all Ni:Cu ratios, not removing any outliers. This is consistent with a relatively small value of EI (0.086—see Table 2).
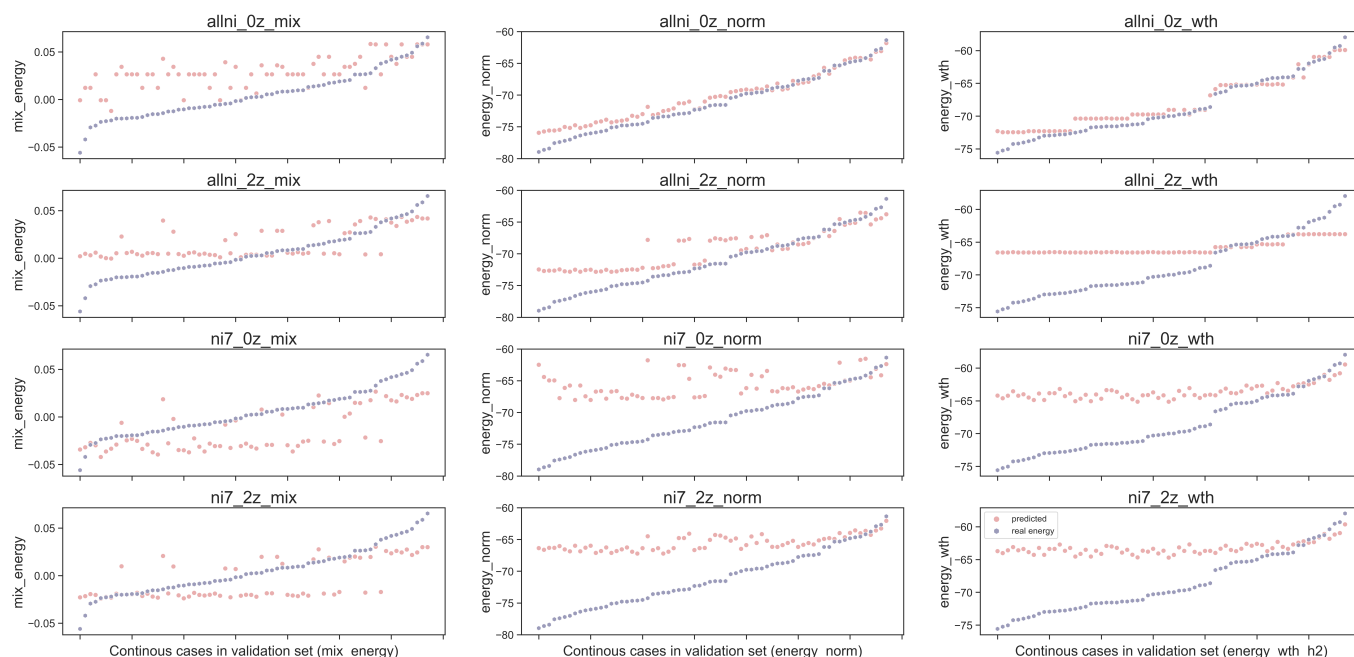
Figure 4. Summary of the accuracy of all models presented in the Table 2. Purple color indicates values from the validation set and pinkish color indicates values predicted by the model.

For this case, it is already possible to indicate a selection of clusters for further analysis with satisfactory accuracy. However, in Section 5.2, we present a slightly more complex approach that leads to much better results.

The results for all trained models are shown in Figure 4. Two main "behaviors" in predicting the energy of complexes can be distinguished: those that contain continuous or discrete changes in values. The number of unique values in the set of values predicted by the model is summarized in Table 3. Interestingly, in most combinations, it was the kernel

models that proved the most successful in predicting the energy of complexes. In the case of ensemble models, the number of unique values dropped significantly. We suspect that this is due to the rather discrete nature of the descriptors—the descriptor characterized by the greatest continuity of values is the distance of the center of mass from the geometric center. There were only 150 unique values of this descriptor in the 650-element set that was used to train the model.

Table 3. Number of unique values predicted by the model; set size = 68.

Model ID	Number of Unique Values	Method	Parameters of Models
allni_0z_norm	55	krr	α: 0.2, kernel: linear
allni_2z_mix	55	krr	α: 3.0, kernel: linear
allni_2z_norm	55	svr	C: 0.2, eps: 0.1, kernel: linear
ni7_0z_mix	55	svr	C: 0.2, eps: 0.1, kernel: linear
ni7_0z_norm	55	krr	α: 1.0, degree: 2, eps: 0.5, kernel: poly
ni7_2z_wth	55	krr	α: 4.0, kernel: sigmoid
ni7_2z_mix	55	krr	α: 3.0, degree: 1, kernel: rbf
ni7_2z_norm	55	krr	α: 1.5, kernel: sigmoid
ni7_0z_wth	55	krr	α: 2.0, kernel: sigmoid
allni_0z_wth	16	ada	est: 25, learning rate: 0.2 loss: linear
allni_0z_mix	13	etr	est: 25, min_sample_split: 2, max_depth: 2
allni_2z_wth	8	rf	est: 25, minimum_sample_split: 2, max_depth: 2

5.2. Conversion from Mixing Energy to Potential Energy

Finally, we decided to test the prediction potential of the trained model to predict mix_energy and subsequently transform the prediction result back to energy_norm. As it was described above, we assumed that the energy mixing procedure "denoises" the information that is more valuable and, as a consequence, it makes it more accessible for the model. Our assumption turned out to be correct, which can be seen in Table 4 for the following pairs: mixing energy and predicted mixing energy (mix–mix), potential energy and predicted potential energy (norm–norm), and potential energy and predicted mixing energy transformed to potential energy (norm–t(mix)). In addition to EI, we included in the table the value of the average prediction error, expressed in eV. The relevant plots of the predicted values are shown in Figure 5.

It can be observed that the EI values for the directly predicted energy values are in almost all cases an order of magnitude higher than for the mixing energy converted back to potential energy. The match between the DFT energy and energy predicted using this method is extraordinary and amounts to 0.2 eV. For example, similar accuracy in the validation set was achieved by Modee et al. [28], who used neural networks to predict the energies of monometallic gallium clusters. Also, this is very close to the accuracy of the DFT method itself, which—depending on the functional used—is on the order of 0.1 eV [56].

Table 4. EI values for mixing energy and predicted mixing energy (mix–mix), potential energy and predicted potential energy (norm–norm), and potential energy and predicted mixing energy transformed to potential energy (norm–t(mix)).

Model ID	Mix-Mix	Norm-Norm	Norm-t(mix)	Avg Error (eV)
allni_0z	0.188	0.108	0.011	0.44
allni_2z	0.202	0.179	0.016	0.45
ni7_0z	0.203	0.088	0.010	0.28
ni7_2z	0.179	0.136	0.007	0.20
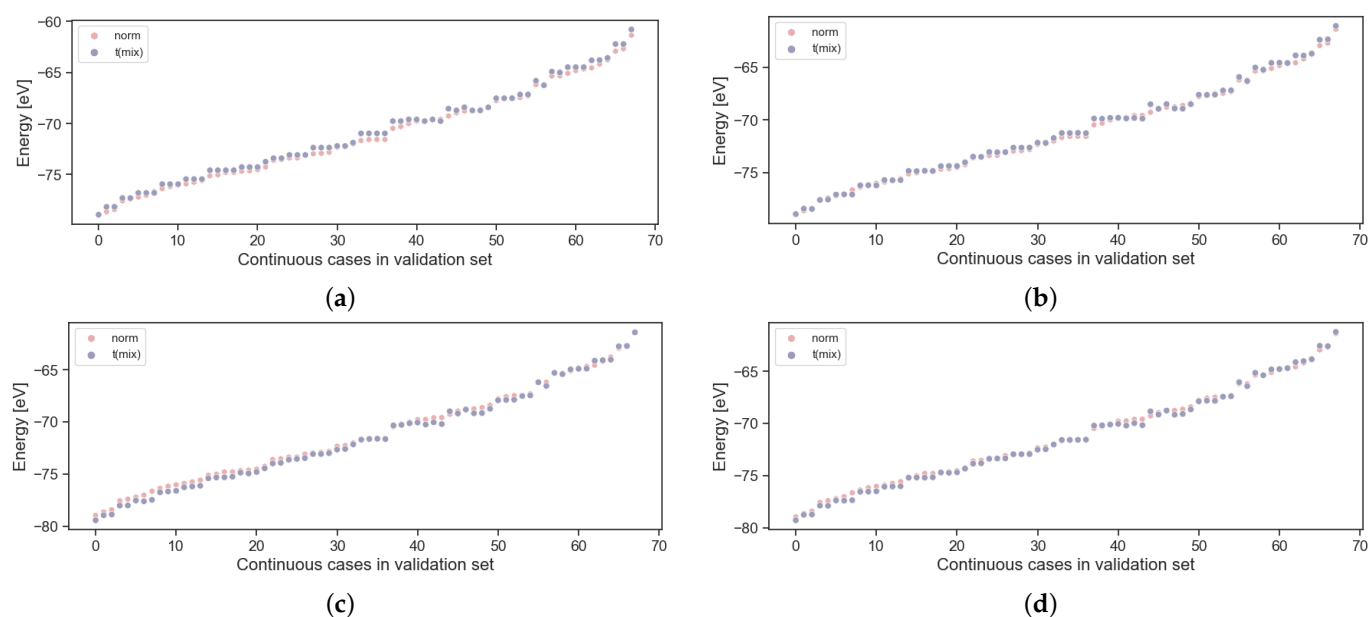
Figure 5. Predicted norm (mix_energy) compared with the validation set on different training settings: (**a**) allni_0z, (**b**) allni_2z, (**c**) ni7_0z, (**d**) ni7_2z.

Expectedly, the Mean Absolute Error (MAE) metric, as seen in Table 2, is not well suited to analyze the quality of the predictions made by our algorithm because of differences in the predicted absolute values. The variation in the mixing energies is much smaller, because this descriptor accounts for the differences in the composition of the cluster; therefore, the MAE for these predictions is much smaller. The MAE for the predicted potential energy values is one or two orders of magnitude greater than for the predicted mixing energy values, which makes a comparison extremely difficult. In general, significantly higher values of MAE for the directly predicted potential energy are in agreement with the more difficult prediction of these values by our algorithm.

5.3. Importance of the Descriptors

Since information about the importance of descriptors is trivial to read from ensemble methods, and in kernel methods it is a value that can be interpreted only with a linear kernel, we decided to only analyze the importance of descriptors in the former group. The results are presented on the heatmap in Figure 6. Each column contains the average importance values for all the tested models (*overall*), trained with the AdaBoost (ada), Random Forest (rf), and Extra Trees Regressor (etr) algorithms. In the subsequent columns, there are values for all the models trained to predict mixing energy (mix), energy corrected for hydrogen (wth), and potential energy (norm). They are followed by all the models trained appropriately without removing outliers (0z) and with the removal of values exceeding the condition Z–score > 2 (2z). The last two columns contain the values for all models trained on sets containing cases with any number of nickel atoms (all Ni) and those restricted to cases with fewer than 7 nickel atoms (Ni < 7).

It can be seen that the importance of information about the atomic composition of the cluster is greater for predicting the energy corrected for hydrogen (*wth* column) and the potential energy (*norm* column) than for the mixing energy (*mix* column). Similarly, the importance of the information on whether a given complex contains a hydrogen atom in its structure is far greater for predicting potential energy than for the energies where corrections for hydrogen were applied (*mix* and *wth* columns). Importantly, in no cases are these features either irrelevant or crucial, which was our main motivation for using different transformations for predicted values instead of simply removing specific descriptors from the training set.

Figure 6. Importance of features for different ensemble methods. Significance values are colored in the following way: the most relevant descriptor for the training set in a given row is the most red, and the least is the most blue.

Surprisingly, in the case of mixing energy, the descriptor with the highest relevance after this transformation was the one about what kind of atom is located in the center of the metallic cluster. This value is very high (0.67 for the mixing energy), while the second most significant descriptor (number of Ni atoms) is assigned an importance level of 0.09. Importantly, the values in the *norm* and *wth* columns show that the information about the central atom is almost ignored by all models. This means that this descriptor is actually significant only for the prediction of mixing energy. In all other cases, this descriptor does not exceed 0.02. It should be noted that in other columns, the importance of the information about the central atom in the cluster is high, but this situation is a consequence of the unusually high importance of this parameter for predicting mixing energy. We conclude that the reason for such a significant increase in the relevance of information about the type of central metal for the mixing energy is the fact that the central metal interacts with every other metal atom in the cluster, and its type has a significant impact on the behavior of the whole system.

Another descriptor that proved important for most of the model training approaches used was the number of Ni–Ni and Cu–Cu bonds in the cluster. Interestingly, information on the number of Ni–Cu bonds was not specifically used by any model. We conclude that the reason for this effect is the strong correlation (positive or negative) of Cu–Cu and Ni–Ni with other parameters (the number of nickel atoms, the composition of the coordination sphere, or information about the central metal), while the Cu–Ni descriptor shows no such correlations. Also, the correlation between these parameters (Cu–Cu and Ni–Ni) alone is R = −0.82, while the correlations of the Cu–Cu and Ni–Ni parameters with Cu–Ni are R = −0.05 and R = −0.53, respectively. Besides, the Cu–Cu and Ni–Ni descriptors are strongly correlated with the predicted values, while the Cu–Ni descriptor does not show similar correlations.

The composition of the coordination sphere may also seem unimportant. However, it should be noted that this is a descriptor composed of several minor descriptors, in which the sum of significance ranges from 0.04 to 0.15, with a mean value of 0.11.

5.4. Predicted Paths of Electrocatalytic CO_2 Reduction

Using the model we trained, it is possible to make a prediction of the energy of a cluster of any geometry and Cu:Ni ratio to which any intermediate considered in Figure 1 is bound. The most accurate model (Table 4) was used to predict potential energies and construct the electrocatalytic plots.

Figure 7 shows the predicted potential energy profiles for the electrocatalytic conversion of CO_2 for formic acid (FA). This process can lead via formate or carboxyl intermediate. Out of the two, the former is more stable. This is because it interacts with the cluster by means of two oxygen bridges, while carboxyl only binds via metal–carbon and metal–oxygen bonds.
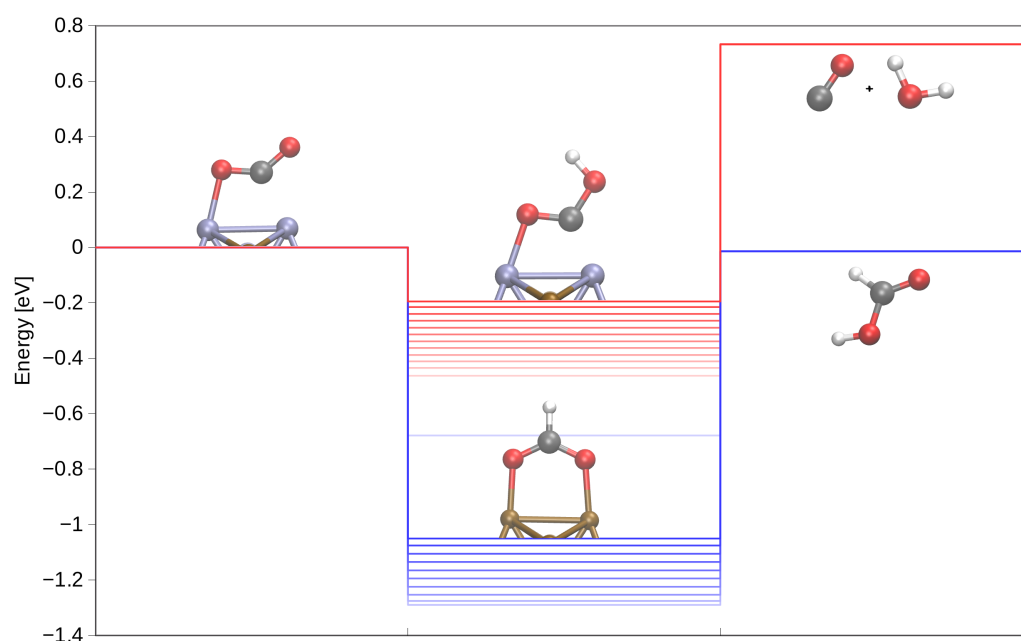
Figure 7. Predicted potential energy profiles for the carboxyl (red) and formate (blue) pathways. The intensity of the color represents the composition of the nanocluster—the darker the color, the greater the Ni content.

Formate is formally negatively charged, but the metallic cluster is able to donate the electron. This implies the strength of the metal–oxygen bonds is the dominating effect in the stabilization of this intermediate. The difference in the potential energy between the formate and carboxyl amounts to approximately 0.7–0.8 eV, regardless of the composition of the cluster.

The subsequent step is a conversion of formate or carboxyl to formic acid. Because of the greater stability of the formate, this step is more difficult and requires greater potential to occur. Contrary to that, the carboxyl-to-formate conversion is easy, and depending on the composition of the cluster, requires from 0.2 to 0.5 eV of external potential. This implies that the external potential can be used to control the selectivity of the process. This observation is consistent with the literature available for pure Cu systems [57], where the applied potential has been determined to influence selectivity. In addition, in the work of Zhao et al. [58], the activation of the formate pathway by alloying Ni with other metals was reported, which further supports our findings.

Alternatively, carboxyl can be further converted to CO with the release of a water molecule. This is not possible to achieve in a single step for the conversion of the formate, as there is already a C-H bond that would have to be cleaved. The carboxyl to CO conversion is endothermic by approximately 1 eV. This results from the assumption made in the present work, where the catalytic reduction products are dissociated from the cluster—our procedure did not involve the prediction of the potential energy of CO bound to the cluster.

Figure 7 also shows that the main factor contributing to the stability of the formate and carboxyl intermediates is the composition of the cluster. We observed a continuous increase in the stability of the intermediate along with an increase in the number of Ni atoms in the cluster. This is also consistent with the literature reports, which show a stronger interaction of carbonyl with Ni than Cu surface [59].

6. Conclusions

We presented a procedure for predicting the energies of Cu_nNi_{13-n} complexes that allows us to determine the energies of these structures without optimization with an average error of 0.2–0.4 eV, which we find surprisingly good for the simplicity of the techniques and the range of simplifications used. We achieved this using only geometric descriptors and relatively simple machine learning methods, which consequently translates into significant time savings compared with optimizing all the structures.

Among the investigated procedures, the one involving the training of the model on the mixing energy and subsequent conversion of the result to potential energy provided the best results.

What we found most unexpected was that the descriptor relevance analysis showed that information about how the metals were distributed in the cluster was not significant in predicting energy; rather, it is the composition of the cluster that is responsible for the stability of the intermediates. However, we believe that this can be easily explained by the d-band of the metallic cluster, which is only partially occupied. This implies that the electron transfer between the particular sites is easy, and it is actually the fundamental reason for the good electric conductance of metals. In the context of our work, it has been confirmed that the number of electrons in the system, i.e., the composition of the cluster, is of much greater importance.

An interesting follow-up study would be to test the transposability of our presented approach to metallic clusters with different compositions or other geometries. The method developed by us obviously can be directly ported only to other bimetallic clusters with icosahedral symmetry. However, in this work, we propose which descriptors are of importance and could be used in the training of machine learning algorithms for other systems. Some of them—such as the composition of the cluster—were trivial to identify, while others—such as the central atom of the cluster— were surprising.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kabir, M.; Habiba, U.E.; Khan, W.; Shah, A.; Rahim, S.; los Rios-Escalante, P.R.D.; Farooqi, Z.U.R.; Ali, L.; Shafiq, M. Climate change due to increasing concentration of carbon dioxide and its impacts on environment in 21st century; a mini review. *J. King Saud Univ. Sci.* **2023**, *35*, 102693. [CrossRef]
2. NOAA's Global Monitoring Lab. Climate Change: Atmospheric Carbon Dioxide. 2022. Available online: https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide (accessed on 25 November 2023).
3. The Intergovernmental Panel on Climate Change. Global Warming of 1.5 °C. 2022. Available online: https://www.ipcc.ch/site/assets/uploads/sites/2/2022/06/SR15_Chapter_1_HR.pdf (accessed on 25 November 2023).
4. Glavič, P.; Pintarič, Z.N.; Levičnik, H.; Dragojlović, V.; Bogataj, M. Transitioning towards Net-Zero Emissions in Chemical and Process Industries: A Holistic Perspective. *Processes* **2023**, *11*, 2647. [CrossRef]
5. Ruscic, B.; Pinzon, R.E.; Morton, M.L.; von Laszevski, G.; Bittner, S.J.; Nijsure, S.G.; Amin, K.A.; Minkoff, M.; Wagner, A.F. Introduction to Active Thermochemical Tables: Several "Key" Enthalpies of Formation Revisited. *J. Phys. Chem. A* **2004**, *108*, 9979–9997. [CrossRef]
6. Zhang, X.; Guo, S.X.; Gandionco, K.A.; Bond, A.M.; Zhang, J. Electrocatalytic carbon dioxide reduction: From fundamental principles to catalyst design. *Mater. Today Adv.* **2020**, *7*, 100074. [CrossRef]
7. Peterson, A.A.; Abild-Pedersen, F.; Studt, F.; Rossmeisl, J.; Nørskov, J.K. How copper catalyzes the electroreduction of carbon dioxide into hydrocarbon fuels. *Energy Environ. Sci.* **2010**, *3*, 1311. [CrossRef]
8. Hori, Y.; Wakebe, H.; Tsukamoto, T.; Koga, O. Electrocatalytic process of CO selectivity in electrochemical reduction of CO_2 at metal electrodes in aqueous media. *Electrochim. Acta* **1994**, *39*, 1833–1839. [CrossRef]
9. DeWulf, D.W.; Jin, T.; Bard, A.J. Electrochemical and Surface Studies of Carbon Dioxide Reduction to Methane and Ethylene at Copper Electrodes in Aqueous Solutions. *J. Electrochem. Soc.* **1989**, *136*, 1686–1691. [CrossRef]
10. Roberts, F.S.; Kuhl, K.P.; Nilsson, A. High Selectivity for Ethylene from Carbon Dioxide Reduction over Copper Nanocube Electrocatalysts. *Angew. Chem. Int. Ed.* **2015**, *54*, 5179–5182. [CrossRef] [PubMed]
11. Hori, Y.; Kikuchi, K.; Murata, A.; Suzuki, S. Production of Methane and Ethylene in Electrochemical Reduction of Carbon Dioxide at Copper Electrode in Aqueous Hydrogencarbonate Solution. *Chem. Lett.* **1986**, *15*, 897–898. [CrossRef]
12. Kim, J.; Summers, D.; Frese, K. Reduction of CO_2 and CO to methane on Cu foil electrodes. *J. Electroanal. Chem. Interfacial Electrochem.* **1988**, *245*, 223–244. [CrossRef]
13. Serafini, M.; Mariani, F.; Fasolini, A.; Scavetta, E.; Basile, F.; Tonelli, D. Nanostructured Copper-Based Electrodes Electrochemically Synthesized on a Carbonaceous Gas Diffusion Membrane with Catalytic Activity for the Electroreduction of CO_2. *ACS Appl. Mater. Interfaces* **2021**, *13*, 57451–57461. [CrossRef]
14. Dongare, S.; Singh, N.; Bhunia, H. Electrocatalytic reduction of CO2 to useful chemicals on copper nanoparticles. *Appl. Surf. Sci.* **2021**, *537*, 148020. [CrossRef]
15. Petranovskii, V.; Gurin, V.; Machorro, R. Spectroscopic observation and ab initio simulation of copper clusters in zeolites. *Catal. Today* **2005**, *107–108*, 892–900. [CrossRef]
16. Piotrowski, M.J.; Piquini, P.; Odashima, M.M.; Silva, J.L.F.D. Transition-metal 13-atom clusters assessed with solid and surface-biased functionals. *J. Chem. Phys.* **2011**, *134*, 134105. [CrossRef]
17. Xiao-Dong, P.; Zhi-Gang, G.; Gong-Ping, L. Energy and structure of copper clusters (n = 70–150) studied by the Monte Carlo computer simulation. *Chin. Phys. B* **2008**, *17*, 3329–3335. [CrossRef]
18. Vilar-Vidal, N.; Rivas, J.; López-Quintela, M.A. Size Dependent Catalytic Activity of Reusable Subnanometer Copper(0) Clusters. *ACS Catal.* **2012**, *2*, 1693–1697. [CrossRef]
19. Grigoryan, V.G.; Alamanova, D.; Springborg, M. Structure and energetics of Cu_N clusters with ($2 \leq n \leq 150$): An embedded-atom-method study. *Phys. Rev. B* **2006**, *73*. [CrossRef]
20. Böyükata, M.; Belchior, J.C. Structural and energetic analysis of copper clusters: MD study of Cu n (n = 2–45). *J. Braz. Chem. Soc.* **2008**, *19*, 884–893. [CrossRef]
21. Alamanova, D.; Grigoryan, V.G.; Springborg, M. Formation of stable products from cluster–cluster collisions. *J. Phys. Condens. Matter* **2007**, *19*, 346204. [CrossRef]
22. Sahoo, S.; Rollmann, G.; Entel, P. Segregation and ordering in binary transition metal clusters. *Phase Transitions* **2006**, *79*, 693–700. [CrossRef]
23. Kyriakou, G.; Boucher, M.B.; Jewell, A.D.; Lewis, E.A.; Lawton, T.J.; Baber, A.E.; Tierney, H.L.; Flytzani-Stephanopoulos, M.; Sykes, E.C.H. Isolated Metal Atom Geometries as a Strategy for Selective Heterogeneous Hydrogenations. *Science* **2012**, *335*, 1209–1212. [CrossRef]
24. Logsdail, A.; Paz-Borbón, L.O.; Johnston, R.L. Structures and Stabilities of Platinum–Gold Nanoclusters. *J. Comput. Theor. Nanosci.* **2009**, *6*, 857–866. [CrossRef]
25. Shin, D.Y.; Won, J.S.; Kwon, J.A.; Kim, M.S.; Lim, D.H. First-principles study of copper nanoclusters for enhanced electrochemical CO2 reduction to CH4. *Comput. Theor. Chem.* **2017**, *1120*, 84–90. [CrossRef]
26. 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. [CrossRef]

27. Janet, J.P.; Kulik, H.J. Predicting electronic structure properties of transition metal complexes with neural networks. *Chem. Sci.* **2017**, *8*, 5137–5152. [CrossRef]

28. Modee, R.; Agarwal, S.; Verma, A.; Joshi, K.; Priyakumar, U.D. DART: Deep learning enabled topological interaction model for energy prediction of metal clusters and its application in identifying unique low energy isomers. *Phys. Chem. Chem. Phys.* **2021**, *23*, 21995–22003. [CrossRef]

29. Zeng, Y.; Li, Q.; Bai, K. Prediction of interstitial diffusion activation energies of nitrogen, oxygen, boron and carbon in bcc, fcc, and hcp metals using machine learning. *Comput. Mater. Sci.* **2018**, *144*, 232–247. [CrossRef]

30. Toyao, T.; Suzuki, K.; Kikuchi, S.; Takakusagi, S.; ichi Shimizu, K.; Takigawa, I. Toward Effective Utilization of Methane: Machine Learning Prediction of Adsorption Energies on Metal Alloys. *J. Phys. Chem. C* **2018**, *122*, 8315–8326. [CrossRef]

31. Omata, K. Screening of New Additives of Active-Carbon-Supported Heteropoly Acid Catalyst for Friedel–Crafts Reaction by Gaussian Process Regression. *Ind. Eng. Chem. Res.* **2011**, *50*, 10948–10954. [CrossRef]

32. Seko, A.; Maekawa, T.; Tsuda, K.; Tanaka, I. Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids. *Phys. Rev. B* **2014**, *89*. [CrossRef]

33. Saad, Y.; Gao, D.; Ngo, T.; Bobbitt, S.; Chelikowsky, J.R.; Andreoni, W. Data mining for materials: Computational experiments with AB compounds. *Phys. Rev. B* **2012**, *85*. [CrossRef]

34. Fletcher, T.L.; Davie, S.J.; Popelier, P.L.A. Prediction of Intramolecular Polarization of Aromatic Amino Acids Using Kriging Machine Learning. *J. Chem. Theory Comput.* **2014**, *10*, 3708–3719. [CrossRef] [PubMed]

35. Snyder, J.C.; Rupp, M.; Hansen, K.; Müller, K.R.; Burke, K. Finding Density Functionals with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 253002. [CrossRef] [PubMed]

36. Balabin, R.M.; Lomakina, E.I. Support vector machine regression (LS-SVM)—an alternative to artificial neural networks (ANNs) for the analysis of quantum chemistry data? *Phys. Chem. Chem. Phys.* **2011**, *13*, 11710. [CrossRef] [PubMed]

37. Faber, F.A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S.S.; Dahl, G.E.; Vinyals, O.; Kearnes, S.; Riley, P.F.; von Lilienfeld, O.A. Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *J. Chem. Theory Comput.* **2017**, *13*, 5255–5264. [CrossRef] [PubMed]

38. Kranz, J.J.; Kubillus, M.; Ramakrishnan, R.; von Lilienfeld, O.A.; Elstner, M. Generalized Density-Functional Tight-Binding Repulsive Potentials from Unsupervised Machine Learning. *J. Chem. Theory Comput.* **2018**, *14*, 2341–2352. [CrossRef] [PubMed]

39. Smith, J.S.; Isayev, O.; Roitberg, A.E. ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192–3203. [CrossRef]

40. Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O.A.; Tkatchenko, A.; Müller, K.R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419. [CrossRef]

41. Rupp, M.; Tkatchenko, A.; Müller, K.R.; von Lilienfeld, O.A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*. [CrossRef]

42. Ma, X.; Li, Z.; Achenie, L.E.K.; Xin, H. Machine-Learning-Augmented Chemisorption Model for CO_2 Electroreduction Catalyst Screening. *J. Phys. Chem. Lett.* **2015**, *6*, 3528–3533. [CrossRef]

43. 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. [CrossRef]

44. Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B Condens. Matter Mater. Phys.* **1999**, *59*, 1758–1775. [CrossRef]

45. Perdew, J.P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868. [CrossRef]

46. Perdew, J.P.; Burke, K.; Ernzerhof, M. Erratum: Generalized gradient approximation made simple (Physical Review Letters (1996) 77 (3865)). *Phys. Rev. Lett.* **1997**, *78*, 1396. [CrossRef]

47. Blöchl, P.E. Projector augmented-wave method. *Phys. Rev. B* **1994**, *50*, 17953–17979. [CrossRef]

48. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

49. Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.; Vapnik, V. Support Vector Regression Machines. In *Advances In Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 1996; Volume 9, pp. 155–161.

50. Hoerl, A.E.; Kennard, R.W. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* **1970**, *12*, 69–82. [CrossRef]

51. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

52. Ho, T.K. Random Decision Forests. In Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1)—Volume 1, Washington, DC, USA, 14–15 August 1995; ICDAR '95, pp. 278–282.

53. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]

54. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]

55. Allen, R.E.; Lidström, S. Life, the universe, and everything—42 fundamental questions. *arXiv* **2018**, arXiv:1804.08730.

56. Bogojeski, M.; Vogt-Maranto, L.; Tuckerman, M.E.; Müller, K.R.; Burke, K. Quantum chemical accuracy from density functional approximations via machine learning. *Nat. Commun.* **2020**, *11*, 5223. [CrossRef] [PubMed]

57. Kortlever, R.; Shen, J.; Schouten, K.J.P.; Calle-Vallejo, F.; Koper, M.T.M. Catalysts and Reaction Pathways for the Electrochemical Reduction of Carbon Dioxide. *J. Phys. Chem. Lett.* **2015**, *6*, 4073–4082. [CrossRef]

58. Zhao, Z.; Chen, Z.; Lu, G. Computational Discovery of Nickel-Based Catalysts for CO2 Reduction to Formic Acid. *J. Phys. Chem. C* **2017**, *121*, 20865–20870. [CrossRef]

59. Liu, C.; Cundari, T.R.; Wilson, A.K. CO_2 Reduction on Transition Metal (Fe, Co, Ni, and Cu) Surfaces: In Comparison with Homogeneous Catalysis. *J. Phys. Chem. C* **2012**, *116*, 5681–5688. [CrossRef]