

# SchNet-IIA: Potential Energy Surface Fitting by Interatomic Interactions Attention Based on Transfer Learning Analysis

Kai-Le Jiang, Huai-Qian Wang,\* Hui-Fang Li, Shu-Wan Pan, Hao Zheng, Yong-Hang Zhang, and Jia-Ming Zhang



Cite This: *J. Chem. Inf. Model.* 2025, 65, 92–100



Read Online

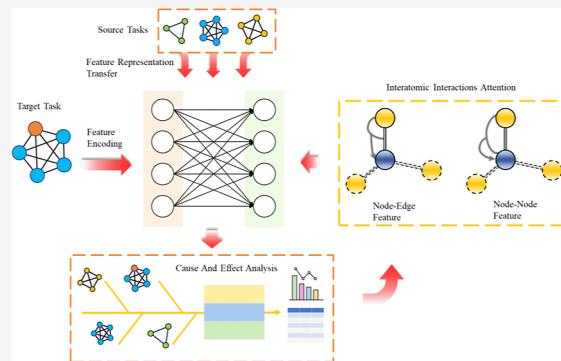
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Machine learning methods for fitting potential energy surfaces and molecular dynamics simulations are becoming increasingly popular due to their potentially high accuracy and savings in computational resources. However, existing application models often rely on basic architectures like artificial neural networks (ANNs) and multilayer perceptron (MLP), lagging behind cutting-edge technologies in the machine learning domain. Furthermore, the complexity of current machine learning frameworks leads to reduced interpretability and challenges for improvement. Herein, we developed a model analysis method based on the feature-representation-transfer approach to directly perform causal analysis on the model. The internal action characteristics of the SchNet framework were successfully analyzed by constructing different source tasks and we proposed interatomic interactions attention for the characterization of doped clusters. The accuracy was enhanced by 0.015 eV/atom compared to the original model. The ability to capture atomic environment characteristics was significantly improved. The activation function was smoothed resulting in a 23.47% increase in the convergence speed. Our SchNet-IIA model demonstrates superior performance in capturing interatomic interactions. Our present work is of distinctive value as it presents a novel transfer learning analysis method with the potential to evolve into a generalized model analysis approach, providing new perspectives and solutions for the field.



## 1. INTRODUCTION

Nanoclusters are extensively studied for their unique microstructures and physical and chemical properties.<sup>1–9</sup> Analyzing the structure and behavior of clusters is of great significance for understanding the properties of materials. Therefore, obtaining these clusters' global minimum (GM) and low-energy structure is essential. However, as the size of the cluster increases, the computing cost required for ab initio calculations increases dramatically due to the size effect.

Recently, data-driven machine learning has become an attractive approach for materials science research.<sup>10–21</sup> Yang et al. combined deep neural networks (DNNs) with transfer learning (TL) for global optimization of Pt clusters and successfully obtained the GM structure of Pt<sub>n</sub> ( $n = 8–14$ ) clusters.<sup>22</sup> Sai et al. developed a cluster graph attention network (CGANet) to globally optimize medium-sized Ag<sub>n</sub> clusters ( $n = 14–26$ ) by aggregating information on adjacent vertices and edges using an attention mechanism.<sup>23</sup> Tkachenko et al. tested the applicability of ANI-1ccx and ANI-nr NN atomic potentials for global minimization of carbon clusters C<sub>n</sub> ( $n = 3–10$ ), and successfully obtained the GM structure of larger clusters such as C<sub>20</sub>.<sup>24</sup> Zhuang and colleagues proposed the concept of first-principles multiscale modeling of mechanical properties based on machine learning interatomic

potential (MLIP).<sup>25,26</sup> In addition, many relatively mature open-source frameworks can be used for potential energy surface fitting or molecular dynamics simulation, such as TorchMD,<sup>27,28</sup> DeePMD-kit,<sup>29,30</sup> and SchNet.<sup>31–35</sup> SchNet is a machine learning architecture that models atomic systems using continuous filter convolutional layers.<sup>34</sup> It follows the Deep Tensor Neural Network (DTNN) framework and builds atomic environments through multilayer interaction blocks.<sup>33</sup>

However, current research of machine learning mainly focuses on monatomic clusters, and there are fewer studies on doped atomic clusters. Since doping atoms introduce complex hybridization effects and more degrees of freedom, it greatly increases the difficulty of searching the potential energy surface of the model.<sup>36</sup> In addition, it is often difficult for multimodule machine learning frameworks to make further improvements due to their complex module interactions.<sup>37,38</sup> To fill this gap, we developed a model analysis method based on transfer

Received: August 16, 2024

Revised: December 22, 2024

Accepted: December 26, 2024

Published: December 30, 2024



learning for improving the interpretability of machine learning models. And we proposed the interatomic interaction attention (IIA) method, in view of the different characteristics brought by doped atoms.

Our overall workflow unfolds as follows. First, we established three data sets, namely  $\text{SeB}_n$ ,  $\text{B}_n$ , and  $\text{SeB}_n\text{--O}$ , for transfer learning and model application. Subsequently, we developed a transfer learning analysis method (TLAM), which was successfully used for the analysis and further optimization of complex machine learning models, effectively enhancing the interpretability of the model. Moreover, based on the results of the analysis for the SchNet framework, an interatomic attention mechanism was proposed to address the complexity of the multiatom cluster system. The SchNet\_IIA model was constructed, and smooth maximum unit was used to smooth the potential energy surface. The results demonstrate that our model has better performance, indicating that it can better capture the interactions between different atoms. This further verifies the effectiveness of the TLAM for model optimization.

## 2. METHODS

**2.1. Data Sets.** Data sets are an essential part of machine learning.<sup>39</sup> In this study, three independent data sets were constructed for model training and transfer learning exploration. The  $\text{SeB}_n$  data set was developed as the training set and the target task, while the  $\text{B}_n$  and  $\text{SeB}_n\text{--O}$  data sets were developed for transfer learning analysis. The  $\text{B}_n$  data set was built and compared with the  $\text{SeB}_n$  data set to analyze the model's ability to capture the characteristics of doped atoms and the atomic environment. The  $\text{SeB}_n\text{--O}$  data set was built and compared with the  $\text{SeB}_n$  data set to analyze the model's sensitivity to high-energy structures and outlier samples. All three data sets used the Artificial Bee Colony (ABC) algorithm<sup>40</sup> to conduct unbiased search for clusters from  $n = 16$  to  $n = 23$  to generate data and combined with DFT calculation at the PBE0/3-21G level.<sup>41</sup> Among them, the conformations in the  $\text{SeB}_n\text{--O}$  data set were globally optimized. All calculations and geometric optimizations in this study were performed using the Gaussian 09 program.<sup>42</sup> In the energy calculations for gas-phase clusters, the contribution of zero-point energy can be considered negligible. Given the small size of the system and its primary focus on chemical bonding interactions, vdW dispersion corrections are considered negligible. The search scale and specific number of data set samples are shown in Table S1.

Since SchNet restricts filters for the continuous-filter convolution layers to be rotationally invariant, there is no need to augment the data set through data augmentation methods, which is also a significant advantage of SchNet's feature prior. During the model training process, the three original data sets were randomly divided in a ratio of 8:1:1, 80% was used as a training set, 10% was used as a validation set, and the remaining 10% was used as a test set alone. We finally developed the SchNet\_IIA model, which integrates the SchNet architecture with IIA modules and SMU, and the optimal configuration hyperparameters of SchNet\_IIA model are shown in bold in Table 1.

**2.2. Transfer Learning.** Transfer learning is based on the basic assumption that there is a certain correlation between different tasks. It transfers the knowledge of a model trained on one task (i.e., the source task) to another related task (i.e., the target task), thereby accelerating and improving the training and performance of the latter.<sup>43</sup> Transfer learning has rapidly

**Table 1. Hyperparameters Searched for the Model**

hyperparameters	test values
$n_{\text{atom\_basis}}^{\text{a}}$	32, 64, <b>128</b> , 256, 512
$n_{\text{interactions}}^{\text{b}}$	2, 3, 4, <b>5</b> , 6, 7, 8, 9, 10
$n_{\text{rbf}}^{\text{c}}$	10, 20, 30, <b>40</b> , 50
cutoff <sup>d</sup>	<b>5</b> , 6, 7, 8, 9
$n_{\text{embedding}}^{\text{e}}$	100, 200, 300, <b>400</b> , 500
learning rate	$10^{-4}$ , $5 \times 10^{-5}$ , <b><math>10^{-5}</math></b> , $5 \times 10^{-6}$

<sup>a</sup>Number of features to describe atomic environments. <sup>b</sup>Number of interaction blocks. <sup>c</sup>Number of radial basis functions. <sup>d</sup>Cutoff radius. <sup>e</sup>Size of the embeddings.

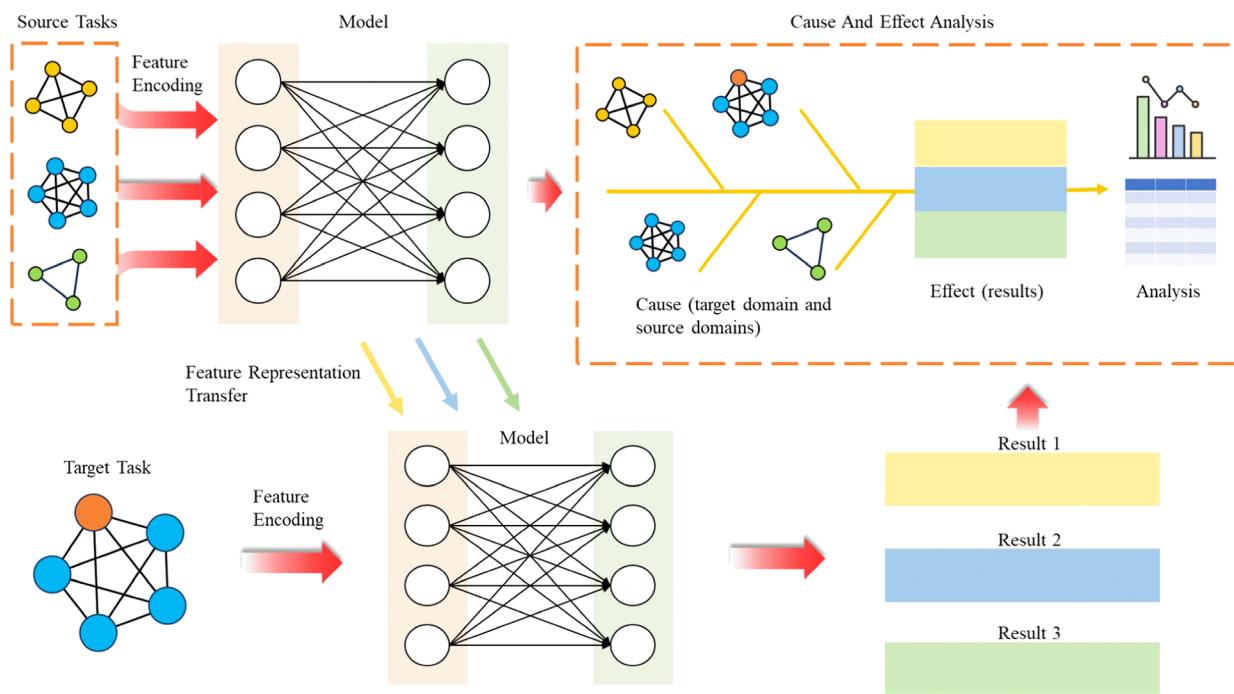
grown in the fields of computer vision and natural language processing (NLP) in recent years,<sup>44,45</sup> and has been applied in the fields of physics and chemistry,<sup>46–48</sup> such as the prediction of polymer-surface adhesion strength.<sup>49</sup> Instance-based, feature-based, parameter-based, and relational-based transfer learning are the four typical types of transfer learning.<sup>50</sup>

When judging which type of transfer learning to use, the following two aspects should be taken into consideration: (a) the main goal that transfer learning needs to accomplish; (b) the degree of correlation between the source task data set and the target task data set. From the perspective of the goals of transfer learning, improving the generalizability of the model, saving computing resources, and improving model accuracy are two common purposes of transfer learning. From the perspective of data set correlation, significant disparities between the element types in the source data set of transfer learning and those in the target learning data set can lead to negative transfer. This phenomenon arises due to insufficient relevance between the source and target domains. This paper adopts feature-based transfer learning to analyze the influence of the connection between feature space and sample space on the model.

**2.3. Transfer Learning Analysis Method.** Neural network models are often viewed as “black box” models.<sup>51,52</sup> Generally, the more complex the model, the less interpretable it is. If the working mechanism inside the neural network cannot be explained, it is difficult to locate the problem.<sup>53</sup> To address the above issues, we developed the TLAM to analyze the physical and chemical meaning of the network in order to further optimize and improve the network structure and obtain a better model. Figure 1 depicts the concept of the entire analysis process.

A domain denoted as  $D = \{X, P(X)\}$ , where  $X$  is the feature space and  $P(X)$  is the marginal distribution probability. Multiple source domains are utilized, which have different feature spaces yet are still related to the target domain, meaning they differ in physical and chemical significance. For different source domains  $D^m$ , they are composed of feature space  $X^m$  and marginal distribution probability  $P(X^m)$ , where  $m$  is the number of the source domain and  $X^m = \{x_1^m, x_2^m, \dots, x_n^m\}$ . Only a single target domain denoted as  $D_T = \{X_T, P(X_T)\}$  is analyzed by this method. For the source domains, we encode features using a mapping function  $f^m: X^m \rightarrow Z$ , where  $Z$  is a common feature space. Finally, the analysis of the regression task is expressed using the difference in performance index values: metric( $g(Z^m)$ ,  $Y_T$ ), where  $g: Z \rightarrow Y$  is the learning model that transfers the common feature space to the label space.

This paper uses TLAM to conduct two different analyses, both with  $\text{SeB}_n$  as the target task. The first use is based on the fact that the mathematical construction of the DTNN model



**Figure 1.** Concept flowchart of transfer learning analysis method.

statistically strictly partitions a wide range of molecular properties into atomic contributions. The source tasks include  $\text{SeB}_{n-1}$ ,  $\text{B}_n$ , and  $\text{B}_{n+1}$ . The chemical meanings of the source domains corresponding to these source tasks relative to the target domains are as follows: (a)  $\text{SeB}_{n-1}$  is doped with a B atom to become  $\text{SeB}_n$  in the transfer process; (b)  $\text{B}_n$  is doped with a Se atom to become  $\text{SeB}_n$  in the transfer process; and (c) one of the B atoms in  $\text{B}_{n+1}$  is replaced by Se to become  $\text{SeB}_n$ . The feature spaces of the target domain and each source domain in Table 2 are first represented as the differences in the

**Table 2. MAE and RMSE of the Original SchNet, as Well as the MAE and RMSE of Transfer Learning Using Different Source Tasks**

	MAE (eV/atom)	RMSE (eV/atom)
the original SchNet	0.118	0.121
$\text{B}_{20}$	0.119	0.121
$\text{B}_{21}$	0.118	0.119
$\text{SeB}_{19}$	0.105	0.105
$\text{SeB}_{19}-\text{O}$	0.110	0.111

kernel charge feature tuples, as shown in Table S2. Based on the results of the first analysis,  $\text{SeB}_{n-1}$  and  $\text{SeB}_{n-1}-\text{O}$  are taken as source tasks in the second time. The difference in features between  $\text{SeB}_{n-1}$  and  $\text{SeB}_{n-1}-\text{O}$  mainly reflects the difference in molecular structure distribution, which leads to the discreteness of molecular energy distribution, as shown in Figure 5 and Figures S1–S7. In the Results and Discussion section, following the TLAM analysis, we employ a continuous transfer learning approach to train the model based on the analysis results, as shown in Figure 2.

The TLAM can be seen as a method that combines experience and causal analysis. It applies known physical and chemical experience to neural networks that lack interpretability. Analyze the causal relationship between the input features of different source domains and the output results of

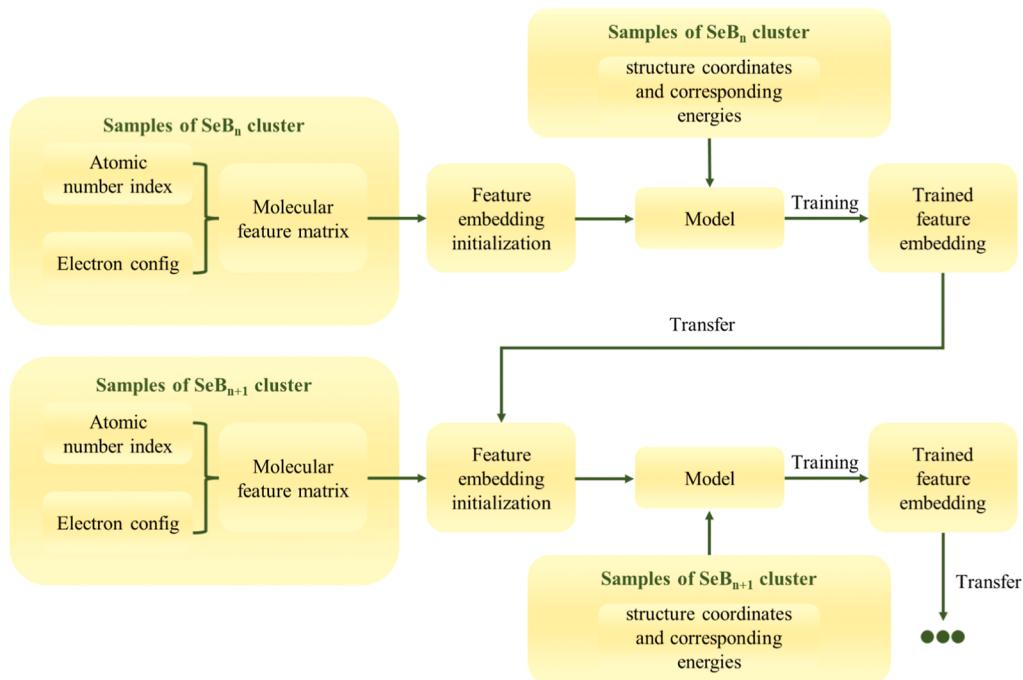
the target domain to explain the source of the prediction results, which increases the interpretability of the model and thus contributes to the optimization and improvement of the model.

#### 2.4. Interatomic Interaction Attention Mechanism.

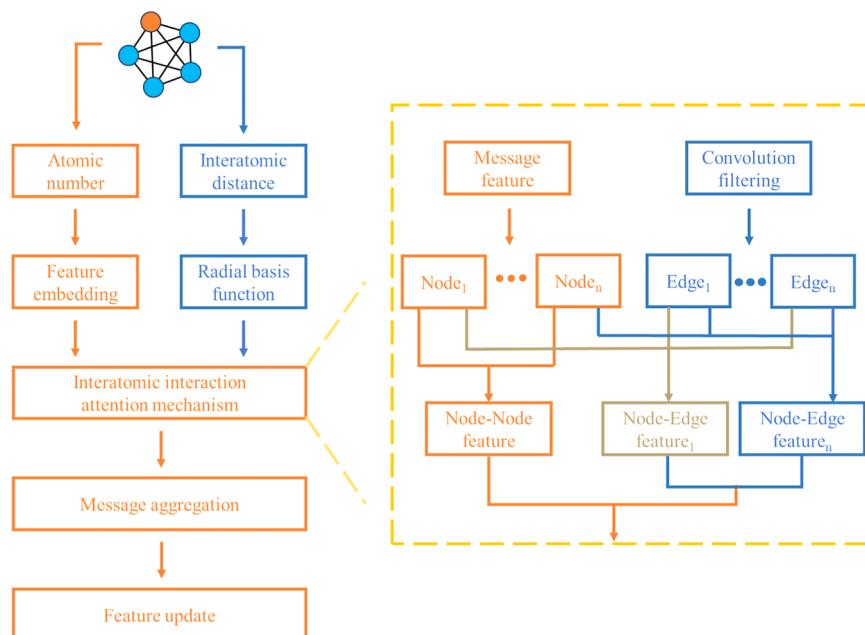
The attention mechanism is a method that determines the importance of each part of the input by learning weights and then redistributes model parameters. It has achieved good performance in modeling molecular properties<sup>54,55</sup> and chemical reactions.<sup>56</sup> In order to better obtain the different characteristics of doped atoms relative to the main atoms, our attention mechanism for atomic interactions is as follows: We construct the characteristic relationship between atoms and bonds by fusing node information with the high-dimensional extended dynamic filtering network of SchNet, while avoiding the computational overhead caused by dense matrix multiplication. Then use the graph attention method to model the importance of interactions between different atoms. Finally, through the idea of residual network, the information on nodes and edges, nodes and nodes is aggregated at the same time, and the characteristic information is updated for output. The model flow is shown in Figure 3. SchNet uses feature tuple  $X = (x_1, x_2, \dots, x_n)$  to represent atoms, where  $n$  is the number of atoms. The filter generation function required by the continuous filter convolution layer is denoted as  $F$ , and the node features output by continuous-filter convolutions are expressed as follows

$$x_i = (X \times F)_i = \sum_j x_j \circ F(r_i - r_j) \quad (1)$$

where “ $\circ$ ” represents the element-wise multiplication, providing input for subsequent aggregation operations, so that the machine learning model can effectively learn the interaction between nodes and edges. Then we further improve on the graph attention mechanism.<sup>59</sup> First, we construct a learnable shared matrix  $W$  to reconstruct the linear transformation of the



**Figure 2.** Continuous transfer learning approach from  $\text{SeB}_n$  to  $\text{SeB}_{n+1}$ .



**Figure 3.** Model flowchart with interatomic interaction attention mechanism.

input node features, and use a feedforward neural network to build a shareable attention function  $f$ . The attention coefficient can be expressed as

$$e_{ij} = f(Wx_i, Wx_j) \quad (2)$$

This is used to describe the importance of the features of node  $j$  to node  $i$ . Then the softmax function is used for normalization and the node features are updated as follows

$$v_i = \sigma(\sum_{k \in N_i} \text{softmax}(e_{ik})Wx_k) \quad (3)$$

$\sigma$  is a nonlinear transformation. The above formula only retains the interaction features between nodes, and discards other structural information, such as edge features. Finally, we aggregate the two features obtained through the idea of residual network and output feature information  $x'_i$

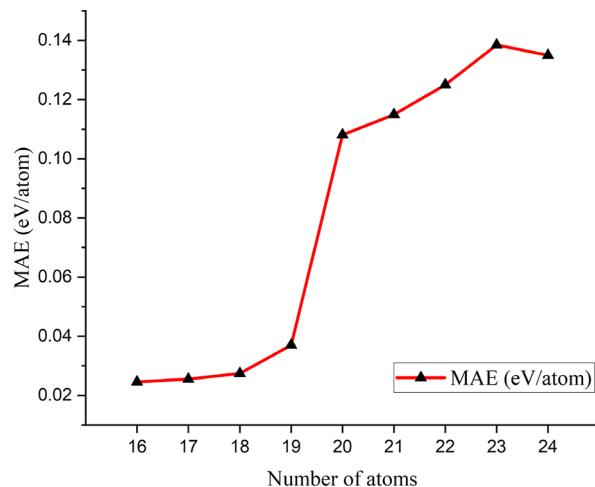
$$x'_i = x_i + v_i \quad (4)$$

Through the attention mechanism based on the interaction between atoms, the model completes the feature learning of nodes and edges, as well as nodes and nodes, more completely captures the atomic environment information, and effectively models the interaction between atoms.

To better provide stability for parameter updates, smooth maximum unit is used to optimize the activation function. SMU is a smooth activation function derived from a smooth approximation of its maximum function. This method uses a smooth approximation of the  $|x|$  function to find a general approximation formula for the maximum function and generalizes to the entire Maxout family.<sup>57</sup>

### 3. RESULT AND DISCUSSION

Figure 4 shows the overall performance captured via mean absolute error (MAE) by the origin SchNet as the size of the



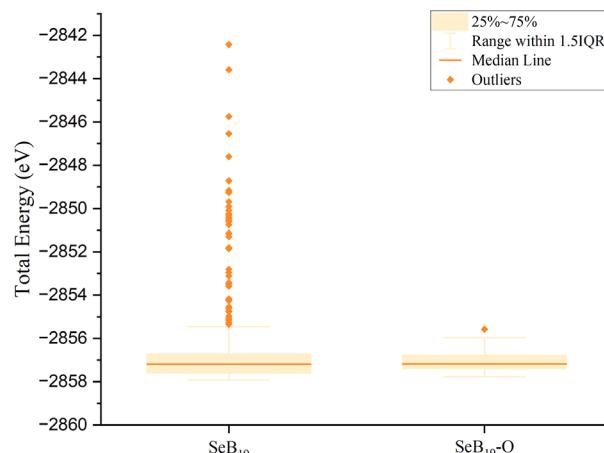
**Figure 4.** MAE results (eV/atom) of clusters of different sizes trained directly with SchNet for  $\text{SeB}_n$ . The  $x$ -axis is the size of the cluster, and the  $y$ -axis is the average MAE of three independent trainings. The final training result represents the MAE value of the 100th epoch.

atomic cluster increases. As the cluster size increases, the potential energy surface fitting becomes more difficult. It can be seen from Figure 4 that from  $\text{SeB}_{19}$  to  $\text{SeB}_{20}$ , the MAE value increases significantly, and the original SchNet begins to have difficulty accurately fitting the potential energy surface of larger-sized clusters. So  $\text{SeB}_{20}$  is taken as a case study in TLAM Results section. We step through the result analysis below.

**3.1. TLAM Results.** Table 2 summarizes the performance of the original SchNet and the  $\text{SeB}_{20}$ 's MAE of transfer using different source tasks. Compared with the original SchNet, the MAEs of the transfer learning models,  $\text{B}_{20}$ ,  $\text{B}_{21}$ , and  $\text{SeB}_{19}$ , are all reduced by 0.119, 0.118, and 0.105 eV/atom, respectively. These demonstrate that transfer learning has benefits for SchNet in fitting energy potential.

Analyzing the different transfer learning tasks of the three source domains  $\text{SeB}_{19}$ ,  $\text{B}_{20}$ , and  $\text{B}_{21}$ , it is obvious that  $\text{SeB}_{19}$  has the best transfer learning effect, and the values of MAE and root mean squared error (RMSE) are 0.105 and 0.105 eV/atom, respectively. From the perspective of atomic composition,  $\text{SeB}_{19}$  contains Se atoms relative to  $\text{B}_{20}$ , and  $\text{B}_{21}$ . After  $\text{SeB}_{19}$  is mapped to the feature space, it provides some feature priors for the feature learning of  $\text{SeB}_{19}$ . This also shows that the SchNet network model focuses on capturing the interactions between different types of atoms, which is consistent with its characteristic of transforming molecular properties into atomic contributions.

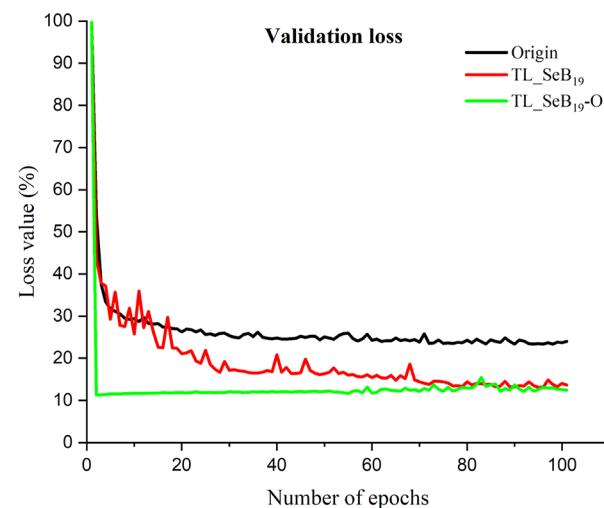
Having examined the local feature learning capabilities of the model, we now evaluate the gradient calculation and error feedback mechanisms for parameter updates. Figure 5 shows



**Figure 5.** Energy distribution of clusters in the  $\text{SeB}_{19}$  and  $\text{SeB}_{19}-\text{O}$  data sets. All conformations in both data sets were calculated by DFT at the PBE0/3-21G level and the conformations in the  $\text{SeB}_{19}-\text{O}$  data set were globally optimized. The box boundaries show the upper quartile ( $Q_3$ ) and the lower quartile ( $Q_1$ ), and the horizontal line inside the box represents the median ( $Q_2$ ). The upper and lower ends of Whiskers extend to  $Q_3 + 1.5 \times \text{IQR}$  and  $Q_1 - 1.5 \times \text{IQR}$ , respectively, where  $\text{IQR} = Q_3 - Q_1$ . The parts beyond the lower quartile or upper quartile are represented by dots, called outliers.

the energy distribution of clusters in the  $\text{SeB}_n$  and  $\text{SeB}_n-\text{O}$  ( $n = 19$ ) data sets. The energy distribution of other clusters is shown in the Figures S1–S7. Through analysis, the molecular relative energy distribution of the  $\text{SeB}_{19}$  data set without structural optimization is more discrete than that of  $\text{SeB}_{19}-\text{O}$ , and contains more high-energy outlier samples. In other words, the  $\text{SeB}_{19}$  data set contains more high-energy structures than that of  $\text{SeB}_{19}-\text{O}$ , which are removed during the opt process due to poor stability and imbalanced molecular configuration. In this case, high-energy structures can be regarded as abnormal samples in machine learning.

Figure 6 depicts the loss curve of  $\text{SeB}_{20}$  transferred learned by  $\text{SeB}_{19}$  and  $\text{SeB}_{19}-\text{O}$ . The epoch when reaching the optimal loss value is shown in the Supporting Information. When the source task is  $\text{SeB}_{19}-\text{O}$ , the first half of the loss curve has a better convergence effect, reaching the lowest loss value at the



**Figure 6.** Validation loss curve of the origin model (black) and transfer learned by  $\text{SeB}_{19}$  (red) and  $\text{SeB}_{19}-\text{O}$  (green).

third epoch, but the second half shows an upward trend and eventually more than the loss value of TL\_SeB<sub>19</sub>. This is due to the fact that there is a lack of high-energy sample features in the source task compared to the target task. The reason for the better initial convergence effect is speculated to be that SeB<sub>19</sub> and SeB<sub>20</sub> have similarities in the low energy part of the potential energy surface. It should be noted that in this part of the transfer learning, when transferring SeB<sub>19</sub> or SeB<sub>19</sub>-O independently to SeB<sub>20</sub>, the computational cost is twice that of direct training at 100 epochs. Further analysis of the remaining clusters detailed in the *Supporting Information* reveals that as the number of outlier samples increases and their distribution becomes more discrete, the convergence capability of the SchNet model gradually weakens. The potential energy surface is typically characterized as smooth and continuous. In contrast, the abnormally high energy samples tend to be discrete and noisy. Given the sensitivity of the SchNet model to such outliers during the overall potential energy surface fitting. This requires a better method to smooth the gradient calculation and optimize the parameter update process.

The TLAM results provide us with directions for model improvement. Through experimental analysis and comparison, SeB<sub>n-1</sub> data set has a better transfer effect than other source data sets. In order to improve the calculation accuracy of the model, the well-trained SeB<sub>16</sub> features are transferred to SeB<sub>17</sub>, and then the well-trained SeB<sub>17</sub> features are transferred to SeB<sub>18</sub>, and this process is repeated continuously to finally complete the model training of SeB<sub>n</sub>. This process is iteratively repeated until the model training for SeB<sub>n</sub> is completed, as illustrated in Figure 2. This transfer learning method reduces the original transfer learning cost from 2 times to 1.2 times, because continuous transfer learning only trained the SeB<sub>16</sub> data set additionally. In order to reduce the number of samples required during the training process of deep neural networks, Yang et al. employed a parameter-based continuous transfer learning method in conjunction with DNNs. This approach successfully facilitated the exploration of Pt clusters.<sup>22</sup>

### 3.2. Ablation Study and Module Improvements

**Results.** The ablation study is designed to evaluate the impact of a certain module or parameter on the performance of the overall model.<sup>58</sup> By systematically removing or modifying specific elements, and observing and analyzing the changes in system behavior, we can ultimately obtain the relative importance of the element and its contribution to the model. This part uses ablation study, and its main purpose is to evaluate the impact of the improved module on fitting accuracy and convergence ability. The SeB<sub>n-1</sub> source domain with the best effect is used for transfer learning in the ablation study. The ablation study results and the overall performance are shown in Table 3. Among them, we designated the overall model as SchNet\_IIA. Its specific components are presented as follows: continuous transfer learning spanning from the SeB<sub>16</sub> data set to the SeB<sub>23</sub> data set, coupled with the addition of interatomic interactions and SMU to the SchNet infrastructure.

The fitting of high-dimensional potential energy surfaces is a prerequisite for cluster structure prediction. However, it is very difficult to accurately describe the potential energy surfaces of polyatomic molecules containing doped atoms because they usually contain multiple minima. Through comparison, the interatomic interaction attention method-based model demonstrated superior performance across accuracy metrics for clusters. In particular, the single application of the IIA module

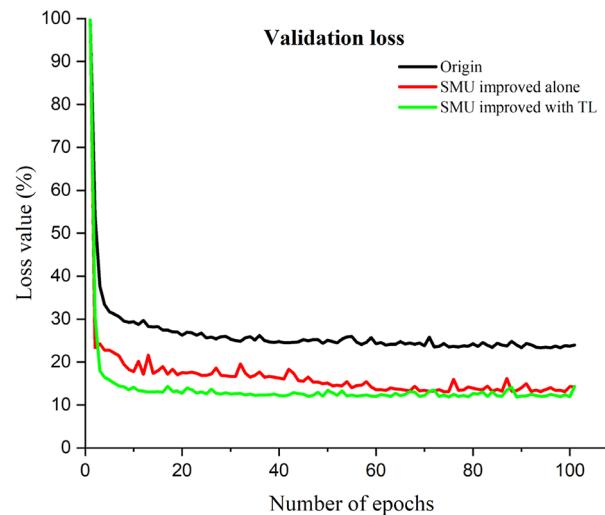
**Table 3. Ablation Study Results**

	MAE (eV/atom)	RMSE (eV/atom)
the original SchNet	0.118	0.121
IIA <sup>a</sup>	0.103	0.103
SMU <sup>b</sup>	0.114	0.116
IIA + TL <sup>c</sup>	0.093	0.094
SMU + TL <sup>d</sup>	0.104	0.105
IIA + SMU <sup>e</sup>	0.098	0.099
SchNet_IIA (IIA + SMU + TL) <sup>f</sup>	0.090	0.090

<sup>a</sup>Interatomic interactions attention mechanism works alone. <sup>b</sup>Smooth maximum unit works alone. <sup>c</sup>Interatomic interactions attention mechanism works with transfer learning. <sup>d</sup>Smooth maximum unit works with transfer learning. <sup>e</sup>Interatomic interactions attention mechanism works with smooth maximum unit. <sup>f</sup>The overall model.

can improve the MAE and RMSE accuracy by 0.015 and 0.018 eV/atom respectively. These results indicate that the interatomic interaction attention method with transfer learning strategy in the SchNet model is a promising approach to accurately capture atomic environment information and fit potential energy surface. Furthermore, the common features Se atoms found between the source domain SeB<sub>n-1</sub> and the target domain SeB<sub>n</sub> represent the shareable knowledge brought by transfer learning. This indicates the model is able to acquire these features by using feature-related data sets.

Figure 7 depicts the loss curves before and after the SMU improved model on the SeB<sub>n</sub> data set. The epoch when



**Figure 7.** Validation loss curves of the original SchNet (black), SMU improved model (red), and SMU with transfer learning joint model (green), respectively.

reaching the optimal loss value is shown in the *Supporting Information*. In terms of convergence of curve, SMU with transfer learning smoothes the loss curve of doped clusters and accelerates convergence to varying degrees. These observed trends are consistent with the TLAM results. Our goal is to obtain a smooth potential energy surface, and it is particularly important to reduce the discontinuity of the activation function. SMU provides much smoother gradients, which helps provide more stable and effective gradient updates during backpropagation, thus the smoothing property of the SMU helps to generate a continuous and smooth potential energy surface. In addition, other systems were used to test the

accuracy of SchNet before and after improvement, and the results are shown in Table S3.

The successful implementation of these methods not only enhances the efficiency and accuracy of the model but also provides valuable insights for future research in the area of fitting potential energy surfaces of complex molecules.

#### 4. CONCLUSION

In this study, we present a novel model analysis method based on feature-representation-transfer approach to conducting causal analysis by transferring different source domains. This approach resolves the issue of difficult optimization due to the interaction of multiple modules in complex models and enhances interpretability. Three independent data sets were constructed starting from the characterization of the chemical significance of the dopant atoms. Through two transfer learning analyses of the SchNet model, it is found that the sensitivity of the model to outliers and verifies that the SchNet model divides molecular properties into atomic contributions. On this basis, we proposed interatomic interactions attention and added SMU method to further optimize the convergence, constructing the SchNet-IIA model. Experimental results show that the SchNet-IIA model further improves the MAE accuracy by 0.028 eV/atom, indicating that the interatomic interaction attention mechanism is a modular tool that effectively captures the characteristics of different atoms. In addition, the parameter update and the robustness of the model have been improved, and the convergence speed has been increased by 23.47%. In the future, the TLAM can be further refined and applied to other machine learning frameworks. Specifically, it is anticipated to be integrated with specific advanced machine learning modules and chemical models to address more challenging problems in the field of potential energy surface fitting and beyond.

#### ■ ASSOCIATED CONTENT

##### Data Availability Statement

All data sets in this work are available for use and viewing in npz format and have been uploaded to GitHub (<https://github.com/Kailejiang/Project-Supplement>). Examples of input files and example scripts for SchNet calculations are available through GitHub and source (<https://github.com/atomistic-machine-learning/schnetpack>).

##### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.4c01473>.

Table S1 presents the search scale and conformations in data sets used for machine learning model analysis and development. Table S2 presents the kernel charge feature tuples of  $B_n$ ,  $B_{n+1}$ ,  $SeB_{n-1}$ ,  $SeB_{n-1}-O$ , and  $SeB_n$  ( $n = 20$ ). Table S3 presents the accuracy of  $Ta_2N_n^-$  ( $n = 3, 5$ ) and  $AlC_n$  ( $n = 21-22$ ) systems before and after model improvement. Table S4 presents the epoch when reaching the optimal loss value of the origin model and transferred learned by  $SeB_n$  and  $SeB_n-O$  in validation. Table S5 presents the epoch when reaching the optimal loss value of the original SchNet, SMU improved model, and SMU with transfer learning joint model (SMU + TL) in validation respectively. Figures S1–S7 show the energy distribution of clusters in the  $SeB_{16}$  and  $SeB_{16}-O$ ,  $SeB_{17}$  and  $SeB_{17}-O$ ,  $SeB_{18}$  and  $SeB_{18}-O$ ,  $SeB_{20}$  and  $SeB_{20}-O$ ,  $SeB_{21}$  and  $SeB_{21}-O$ ,  $SeB_{22}$  and  $SeB_{22}-O$ , and

$SeB_{23}$  and  $SeB_{23}-O$  data sets respectively. All conformations in both data sets were calculated by DFT at the PBE0/3-21G level and the conformations in the  $SeB_n-O$  data set were globally optimized. The box boundaries show the upper quartile (Q3) and the lower quartile (Q1), and the horizontal line inside the box represents the median (Q2). The upper and lower ends of Whiskers extend to  $Q3 + 1.5 \times IQR$  and  $Q1 - 1.5 \times IQR$ , respectively, where  $IQR = Q3 - Q1$ . The parts beyond the lower quartile or upper quartile are represented by dots, called outliers. (PDF)

#### ■ AUTHOR INFORMATION

##### Corresponding Author

Huai-Qian Wang – College of Information Science and Engineering, Huaqiao University, Xiamen 361021, China; College of Engineering, Huaqiao University, Quanzhou 362021, China;  [orcid.org/0000-0003-0388-510X](https://orcid.org/0000-0003-0388-510X); Email: [hqwang@hqu.edu.cn](mailto:hqwang@hqu.edu.cn)

##### Authors

Kai-Le Jiang – College of Information Science and Engineering, Huaqiao University, Xiamen 361021, China  
Hui-Fang Li – College of Engineering, Huaqiao University, Quanzhou 362021, China  
Shu-Wan Pan – College of Engineering, Huaqiao University, Quanzhou 362021, China  
Hao Zheng – College of Information Science and Engineering, Huaqiao University, Xiamen 361021, China  
Yong-Hang Zhang – College of Information Science and Engineering, Huaqiao University, Xiamen 361021, China  
Jia-Ming Zhang – College of Information Science and Engineering, Huaqiao University, Xiamen 361021, China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jcim.4c01473>

##### Notes

The authors declare no competing financial interest.

#### ■ ACKNOWLEDGMENTS

The project was supported by the Natural Science Foundation of Fujian Province of China (Grant no. 2023J01141), the Natural Science Foundation of Xiamen (Grant no. 3502Z202373051).

#### ■ REFERENCES

- (1) Castleman, A. W. From Elements to Clusters: The Periodic Table Revisited. *J. Phys. Chem. Lett.* **2011**, 2 (9), 1062–1069.
- (2) Jamwal, D.; Sharma, A.; Kanwar, R.; Mehta, S. K. The Multifaceted Dimensions of Potent Nanostructures: A Comprehensive Review. *Mater. Chem. Front.* **2021**, 5, 2967–2995.
- (3) Sergeeva, A. P.; Popov, I. A.; Piazza, Z. A.; Li, W. L.; Romanescu, C.; Wang, L. S.; Boldyrev, A. I. Understanding Boron through Size-Selected Clusters: Structure, Chemical Bonding, and Fluxionality. *Acc. Chem. Res.* **2014**, 47 (4), 1349–1358.
- (4) Duan, Q. Z.; Shen, J. Y.; Zhong, X.; Lu, H. Y.; Lu, C. Structural Phase Transition and Superconductivity of Ytterbium under High Pressure. *Phys. Rev. B* **2022**, 105 (21), 214503.
- (5) Ciuparu, D.; Klie, R. F.; Zhu, Y. M.; Pfefferle, L. Synthesis of Pure Boron Single-Wall Nanotubes. *J. Phys. Chem. B* **2004**, 108 (13), 3967–3969.
- (6) Liu, F.; Shen, C. M.; Su, Z. J.; Ding, X. L.; Deng, S. Z.; Chen, J.; Xu, N. S.; Gao, H. J. Metal-like Single Crystalline Boron Nanotubes:

- Synthesis and In Situ Study on Electric Transport and Field Emission Properties. *J. Mater. Chem.* **2010**, *20*, 2197–2205.
- (7) Fan, Y. W.; Wang, H. Q.; Li, H. F. Probing the Structural and Electronic Properties of Anionic Europium-Doped Silicon Clusters by Density Functional Theory and Comparison of Experimental Photoelectron Spectroscopy. *Chem. Phys.* **2020**, *538*, 110918.
- (8) Wang, H. Q.; Li, H. F. Structure Identification of Endohedral Golden Cage Nanoclusters. *RSC Adv.* **2015**, *5*, 94685–94693.
- (9) Li, H. F.; Wang, H. Q. Stabilization of Golden Cages by Encapsulation of a Single Transition Metal Atom. *R. Soc. Open Sci.* **2018**, *5* (1), 171019.
- (10) Behler, J. Constructing High-Dimensional Neural Network Potentials: A Tutorial Review. *Int. J. Quantum Chem.* **2015**, *115*, 1032–1050.
- (11) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), 146401.
- (12) Behler, J. Atom-Centered Symmetry Functions for Constructing High-Dimensional Neural Network Potentials. *J. Chem. Phys.* **2011**, *134* (7), 074106.
- (13) Artrith, N.; Morawietz, T.; Behler, J. High-Dimensional Neural-Network Potentials for Multicomponent Systems: Applications to Zinc Oxide. *Phys. Rev. B* **2011**, *83* (15), 153101.
- (14) Artrith, N.; Hiller, B.; Behler, J. Neural Network Potentials for Metals and Oxides - First Applications to Copper Clusters at Zinc Oxide. *Phys. Status Solidi B* **2013**, *250*, 1191–1203.
- (15) Handley, C. M.; Popelier, P. L. A. Potential Energy Surfaces Fitted by Artificial Neural Networks. *J. Phys. Chem. A* **2010**, *114* (10), 3371–3383.
- (16) Chen, X.; Jørgensen, M. S.; Li, J.; Hammer, B. Atomic Energies from a Convolutional Neural Network. *J. Chem. Theory Comput.* **2018**, *14* (7), 3933–3942.
- (17) Zhai, H.; Alexandrova, A. N. Ensemble-Average Representation of Pt Clusters in Conditions of Catalysis Accessed through GPU Accelerated Deep Neural Network Fitting Global Optimization. *J. Chem. Theory Comput.* **2016**, *12* (12), 6213–6226.
- (18) Khorshidi, A.; Peterson, A. A. Amp A Modular Approach to Machine Learning in Atomistic Simulations. *Comput. Phys. Commun.* **2016**, *207*, 310–324.
- (19) Ouyang, R.; Xie, Y.; Jiang, D. E. Global Minimization of Gold Clusters by Combining Neural Network Potentials and the Basin-Hopping Method. *Nanoscale* **2015**, *7*, 14817–14821.
- (20) Ma, S.; Shang, C.; Liu, Z.-P. Heterogeneous Catalysis from Structure to Activity via SSW-NN Method. *J. Chem. Phys.* **2019**, *151* (5), 050901.
- (21) Han, L.; Jiang, G.-D.; Li, X.-N.; He, S.-G. Global Optimization of Ta<sub>n</sub> Clusters by Deep Neural Network. *Chem. Phys. Lett.* **2021**, *785*, 139118.
- (22) Yang, Q.; Jiang, G.; He, S. Enhancing the Performance of Global Optimization of Platinum Cluster Structures by Transfer Learning in a Deep Neural Network. *J. Chem. Theory Comput.* **2023**, *19* (6), 1922–1930.
- (23) Sai, L.; Fu, L.; Du, Q.; Zhao, J. Graph attention network for global search of atomic clusters: A case study of Ag<sub>n</sub> (n = 14–26) clusters. *Front. Phys.* **2023**, *18*, 13306.
- (24) Tkachenko, N. V.; Tkachenko, A. A.; Nebgen, B.; Tretiak, S.; Boldyrev, A. I. Neural network atomistic potentials for global energy minima search in carbon clusters. *Phys. Chem. Chem. Phys.* **2023**, *25*, 21173–21182.
- (25) Mortazavi, B.; Silani, M.; Podryabinkin, E. V.; Rabczuk, T.; Zhuang, X.; Shapeev, A. V. First-Principles Multiscale Modeling of Mechanical Properties in Graphene/Borophene Heterostructures Empowered by Machine-Learning Interatomic Potentials. *Adv. Mater.* **2021**, *33*, 2102807.
- (26) Mortazavi, B.; Javvaji, B.; Shojaei, F.; Rabczuk, T.; Shapeev, A. V.; Zhuang, X. Exceptional Piezoelectricity High Thermal Conductivity and Stiffness, and Promising Photocatalysis in Two-Dimensional MoSi<sub>2</sub>N<sub>4</sub> Family Confirmed by First-Principles. *Nano Energy* **2021**, *82*, 105716.
- (27) Doerr, S.; Majewski, M.; Pérez, A.; Krämer, A.; Clementi, C.; Noé, F.; Giorgino, T.; De Fabritiis, G. TorchMD: A Deep Learning Framework for Molecular Simulations. *J. Chem. Theory Comput.* **2021**, *17* (4), 2355–2363.
- (28) Pelaez, R. P.; Simeon, G.; Galvelis, R.; Mirarchi, A.; Eastman, P.; Doerr, S.; Thölke, P.; Markland, T. E.; De Fabritiis, G. TorchMD-Net 2.0: Fast Neural Network Potentials for Molecular Simulations. *J. Chem. Theory Comput.* **2024**, *20* (10), 4076–4087.
- (29) Wang, H.; Zhang, L.; Han, J.; E, W. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **2018**, *228*, 178–184.
- (30) Zeng, J.; Zhang, D.; Lu, D.; Mo, P.; Li, Z.; Chen, Y.; Rynik, M.; Huang, L.; Li, Z.; Shi, S.; Wang, Y.; Ye, H.; Tuo, P.; Yang, J.; Ding, Y.; Li, Y.; Tisi, D.; Zeng, Q.; Bao, H.; Xia, Y.; Huang, J.; Muraoka, K.; Wang, Y.; Chang, J.; Yuan, F.; Bore, S. L.; Cai, C.; Lin, Y.; Wang, B.; Xu, J.; Zhu, J.-X.; Luo, C.; Zhang, Y.; Goodall, R. E. A.; Liang, W.; Singh, A. K.; Yao, S.; Zhang, J.; Wentzcovitch, R.; Han, J.; Liu, J.; Jia, W.; York, D. M.; E, W.; Car, R.; Zhang, L.; Wang, H. J. DeePMD-kit v2: A software package for Deep Potential models. *Chem. Phys.* **2023**, *159*, 054801.
- (31) Schütt, K. T.; Hessmann, S. S. P.; Gebauer, N. W.; Lederer, J.; Gastegger, M. SchNetPack 2.0: A neural network toolbox for atomistic machine learning. *J. Chem. Phys.* **2023**, *158* (14), 144801.
- (32) Schütt, K. T.; Kessel, P.; Gastegger, M.; Nicoli, K. A.; Tkatchenko, A.; Müller, K. R. SchNetPack A Deep Learning Toolbox For Atomistic Systems. *J. Chem. Theory Comput.* **2018**, *15* (1), 448–455.
- (33) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **2017**, *8*, 13890.
- (34) Schütt, K. T.; Kindermans, P.-J.; Sauceda, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in Neural Information Processing Systems*; MIT Press 2017; Vol. 30, pp 992–1002.
- (35) Schütt, K. T.; Sauceda, H. E.; Kindermans, P. J.; Tkatchenko, A.; Müller, K. R. SchNet-a deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148* (24), 241722.
- (36) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121* (16), 10142–10186.
- (37) Ahmed, S. F.; Alam, M. S. B.; Hassan, M.; Rozbu, M. R.; Ishtiaq, T.; Rafa, N.; Mofijur, M.; Shawkat Ali, A. B. M.; Gandomi, A. H. Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artif. Intell. Rev.* **2023**, *56* (11), 13521–13617.
- (38) Alzubaidi, L.; Zhang, J.; Humaidi, A. J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M. A.; Al-Amidie, M.; Farhan, L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data.* **2021**, *8*, 53.
- (39) Gong, Y.; Liu, G.; Xue, Y.; Li, R.; Meng, L. A survey on dataset quality in machine learning. *Inform. Softw. Technol.* **2023**, *162*, 107268.
- (40) Zhang, J.; Dolg, M. ABCluster: the artificial bee colony algorithm for cluster global optimization. *Phys. Chem. Chem. Phys.* **2015**, *17*, 24173–24181.
- (41) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.
- (42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo,

J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision C.01; Gaussian Inc.: Wallingford CT, 2010.

(43) Hosna, A.; Merry, E.; Gyalmo, J.; Alom, Z.; Aung, Z.; Azim, M. A. Transfer Learning: A Friendly Introduction. *J. Big Data* **2022**, *9*, 102.

(44) Sung, Y.-L.; Cho, J.; Bansal, M. VL-ADAPTER: Parameter-Efficient Transfer Learning for Vision-and-Language Tasks. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022; IEEE: New Orleans, LA, USA; pp 5217–5227.

(45) Kora, P.; Ooi, C. P.; Faust, O.; Raghavendra, U.; Gudigar, A.; Chan, W. Y.; Meenakshi, K.; Swaraja, K.; Plawiak, P.; Rajendra Acharya, U. Transfer Learning Techniques for Medical Image Analysis: A Review. *Biocybern. Biomed. Eng.* **2022**, *42* (1), 79–107.

(46) Pesciullesi, G.; Schwaller, P.; Laino, T.; Reymond, J.-L. Transfer Learning Enables the Molecular Transformer to Predict Regio- and Stereoselective Reactions on Carbohydrates. *Nat. Commun.* **2020**, *11*, 4874.

(47) Meuwly, M. Machine Learning for Chemical Reactions. *Chem. Rev.* **2021**, *121* (16), 10218–10239.

(48) Cai, C.; Wang, S.; Xu, Y.; Zhang, W.; Tang, K.; Ouyang, Q.; Lai, L.; Pei, J. Transfer Learning for Drug Discovery. *J. Med. Chem.* **2020**, *63* (16), 8683–8694.

(49) Shi, J.; Albreiki, F.; Colón, Y. J.; Srivastava, S.; Whitmer, J. K. Transfer Learning Facilitates the Prediction of Polymer-Surface Adhesion Strength. *J. Chem. Theory Comput.* **2023**, *19* (14), 4631–4640.

(50) Pan, S. J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22* (10), 1345–1359.

(51) Doshi-Velez, F.; Kim, B. Towards A Rigorous Science of Interpretable Machine Learning. *arXiv* **2017**, arXiv: 1702.08608.

(52) Deng, C.; Ji, X.; Rainey, C.; Zhang, J.; Lu, W. Integrating Machine Learning with Human Knowledge. *iScience* **2020**, *23* (11), 101656.

(53) Montavon, G.; Samek, W.; Müller, K.-R. Methods for Interpreting and Understanding Deep Neural Networks. *Digit. Signal Process.* **2018**, *73*, 1–15.

(54) Liu, H.; Huang, Y.; Liu, X.; Deng, L. Attention-wise Masked Graph Contrastive Learning for Predicting Molecular Property. *Brief. Bioinform.* **2022**, *23* (5), bbac303.

(55) Ye, X.-b.; Guan, Q.; Luo, W.; Fang, L.; Lai, Z.-R.; Wang, J. Molecular Substructure Graph Attention Network for Molecular Property Identification in Drug Discovery. *Pattern Recogn.* **2022**, *128*, 108659.

(56) Sacha, M.; Blaž, M.; Byrski, P.; Dąbrowski-Tumanński, P.; Chromiński, M.; Łoska, R.; Włodarczyk-Pruszyński, P.; Jastrzębski, S. Molecule Edit Graph Attention Network: Modeling Chemical Reactions as Sequences of Graph Edits. *J. Chem. Inf. Model.* **2021**, *61* (7), 3273–3284.

(57) Biswas, K.; Kumar, S.; Banerjee, S.; Pandey, A. K. Smooth Maximum Unit: Smooth Activation Function for Deep Networks using Smoothing Maximum Technique. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022, 784–793.

(58) Meyes, R.; Lu, M.; Waubert de Puiseau, C.; Meisen, T. Ablation Studies in Artificial Neural Networks. *arXiv* **2019**, arXiv:1901.08644v2.

(59) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2017**, arXiv:1710.10903v3.



#### CAS INSIGHTS™

## EXPLORE THE INNOVATIONS SHAPING TOMORROW

Discover the latest scientific research and trends with CAS Insights. Subscribe for email updates on new articles, reports, and webinars at the intersection of science and innovation.

[Subscribe today](#)

**CAS**  
A division of the  
American Chemical Society