

ARTICLE OPEN



Predicting adsorption ability of adsorbents at arbitrary sites for pollutants using deep transfer learning

Zhilong Wang^{1,2,4}, Haikuo Zhang^{1,2,4}, Jiahao Ren^{1,2}, Xirong Lin^{1,2}, Tianli Han³, Jinyun Liu³✉ and Jinjin Li^{1,2}✉

Accurately evaluating the adsorption ability of adsorbents for heavy metal ions (HMIs) and organic pollutants in water is critical for the design and preparation of emerging highly efficient adsorbents. However, predicting adsorption capabilities of adsorbents at arbitrary sites is challenging, with currently unavailable measuring technology for active sites and the corresponding activities. Here, we present an efficient artificial intelligence (AI) approach to predict the adsorption ability of adsorbents at arbitrary sites, as a case study of three HMIs (Pb(II), Hg(II), and Cd(II)) adsorbed on the surface of a representative two-dimensional graphitic-C₃N₄. We apply the deep neural network and transfer learning to predict the adsorption capabilities of three HMIs at arbitrary sites, with the predicted results of Cd(II) > Hg(II) > Pb(II) and the root-mean-squared errors less than 0.1 eV. The proposed AI method has the same prediction accuracy as the ab initio DFT calculation, but is millions of times faster than the DFT to predict adsorption abilities at arbitrary sites and only requires one-tenth of datasets compared to training from scratch. We further verify the adsorption capacity of g-C₃N₄ towards HMIs experimentally and obtain results consistent with the AI prediction. It indicates that the presented approach is capable of evaluating the adsorption ability of adsorbents efficiently, and can be further extended to other interdisciplines and industries for the adsorption of harmful elements in aqueous solution.

npj Computational Materials (2021)7:19; <https://doi.org/10.1038/s41524-021-00494-9>

INTRODUCTION

Recent studies have shown that when artificial intelligence (AI) meets material design and discovery, it means reducing the time and cost going from lab to practical applications by greatly improving the research efficiency^{1–3}. Heavy metal ions (HMIs) and organic pollutants are major sources for water pollution^{4,5}, causing persistent harm through the accumulation in food chain, threatening ecological conditions and human health^{6,7}. Pioneers have designed and synthesized several adsorbents that exhibit high adsorption ability for removing HMIs and organic pollutants from water^{8–13}. Adsorption ability of an adsorbent relies on the active sites and the corresponding activity intensities, which is currently hardly detectable^{14,15}. Theoretical prediction provides an alternative approach for understanding the mechanism of the adsorption process and for exploring highly efficient adsorbents. Researchers spend a lot of time to allocate, model, and wait for first-principle calculations^{16–19}, which can determine the adsorption capacity of materials at special sites in advance. However, the configurational space offered by the wide variety of materials and the complex relationships between active sites and activity intensities of adsorbents indicates that a conventional approach for structural optimization, based on inherently time-consuming ab initio methods, is particularly challenging.

Recently, the means that are based on mechanism have been partly displaced by machine learning (ML), which is an AI method containing three elements: models, strategies, and algorithms, so as to speed up the computational process and obtain complex physical and chemical properties that are not accessible with conventional approaches^{20–22}. While significant research progress has been achieved by improving the material descriptors over

many years, the applications of ML for material prediction is in general plagued by several significant challenges^{23–25}. For example, for some tasks, to achieve a high prediction accuracy, the ML method requires a sufficient amount of effective data to capture the correlations between physical properties and features or uses repeated iterations to train different models, which inevitably consumes time and reduces the efficiency of ML^{26–28}. To address those issues, in this study, we present a popular ML model to investigate the HMI trapping and quantitatively determine the adsorption ability of adsorbent to HMIs at arbitrary sites. The transfer learning (TL) method is adopted in the model^{29–32}, which has hardly been mentioned and applied in the adsorption energy prediction model. Since two-dimensional materials commonly possess enriched adsorption active sites at several positions (such as defect and boundary) with abundant surface functional groups, especially, the ultra-thin two-dimensional materials can have large surface area because the material can maintain the maximum plane size while maintaining the atomic thickness^{33–35}, they have been considered promising adsorbents for many fields including water purification. Herein, we choose a typical two-dimensional (2D) graphitic-C₃N₄ (g-C₃N₄) adsorbent as a case study to evaluate the adsorption characteristics towards three representative HMIs including Pb(II), Hg(II), and Cd(II).

Unlike most ML approaches that use different models for training and testing based on enough data to ensure accuracy and avoid overfitting, the TL method can transfer knowledge from one dataset to another in the different but related domains with high reliability, making full use of the feature similarity between models. For the prediction of similar material properties, TL alleviates the issues of time-consuming and data scarcity by

¹National Key Laboratory of Science and Technology on Micro/Nano Fabrication, Shanghai Jiao Tong University, Shanghai, China. ²Key Laboratory for Thin Film and Microfabrication of Ministry of Education, Department of Micro/Nano-electronics, Shanghai Jiao Tong University, Shanghai, China. ³Key Laboratory of Functional Molecular Solids of Ministry of Education, Anhui Provincial Engineering Laboratory for New-Energy Vehicle Battery Energy-Storage Materials, Anhui Laboratory of Molecule-Based Materials, College of Chemistry and Materials Science, Anhui Normal University, Wuhu, Anhui, China. ⁴These authors contributed equally: Zhilong Wang, Haikuo Zhang. ✉email: jyliu@ahnu.edu.cn; lijinjin@sjtu.edu.cn

switching the multi-model training to single-model training, decreasing a large amount of training data to a small amount of effective data. TL is able to utilize the chemical and physical properties and similarities between the structure descriptors learned by Pb(II)/g-C₃N₄ model, as well as Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ models. Based on the TL method, the adsorption ability towards Hg(II), Cd(II) at arbitrary sites can be predicted accurately and quickly by a small amount of training data, through training the adsorption ability of Pb(II) on the surface of g-C₃N₄ in advance.

In our study, 7000 adsorption energies calculated by the ab initio density functional theory (DFT) were used to predict the adsorption sites and adsorption capacity of Pb(II)/g-C₃N₄ through the deep neural network (DNN), which served as the initial model to be learned. Based on the TL method, the adsorption capacity of the remaining HMIs on the same adsorbent can be predicted by a small amount of DFT data. Here, only 700 adsorption energies were calculated to quickly predict the adsorption capacity of Hg(II) and Cd(II) on the surface of g-C₃N₄ through TL. DNN prediction indicated that compared with the edges of adsorbent material, HMIs were more likely to be adsorbed at the center of g-C₃N₄ adsorbent, with predicted RMSEs all less than 0.1 eV. A RMSE of 0.1 eV by a prediction model with only a few hundreds of DFT calculations is treated as a remarkable feat, which provides a powerful guarantee for predicting the adsorption capacity of adsorbent towards HMIs accurately^{27,36}. The presented AI method has the same accuracy as the ab initio DFT calculation, but is ten times faster than the training from scratch in the training stage (only requires one-tenth of datasets than training from scratch) and millions of times faster than the DFT in prediction stage. In addition to the adsorption ability prediction of g-C₃N₄ for Pb(II), Hg(II), and Cd(II), the proposed method can be easily extended to predict the adsorption ability of other adsorbents for different HMIs, organic contaminants, etc., which is significant for the environmental treatment of removing harmful pollutants from water.

RESULTS

Adsorption ability prediction of g-C₃N₄ for Pb(II)

We started with the determination of the adsorption ability of Pb(II) adsorbed at the arbitrary site of the surface of g-C₃N₄. The corresponding adsorption model is presented as Pb(II)/g-C₃N₄. To ensure the unbiased statistical results, a total of 7000 single-point adsorption energies with different potential active sites were calculated by DFT. The Deep Potential-Smooth Edition (DeepPot-SE), an end-to-end deep neural network-based (DNN) potential energy surface (PES) model, was performed to evaluate the adsorption ability to Pb(II) adsorbed on the surface of g-C₃N₄ at arbitrary site in the feature space. Figure 1 shows the schematic model of the adsorption process of HMIs on the surface of g-C₃N₄, where the active sites and activity intensities of Pb(II) adsorbed on the surface of g-C₃N₄ were calculated by DFT and trained by the DNN model, while the corresponding adsorption ability of Hg(II) and Cd(II) can be predicted by a small amount of data via TL method. Supplementary Fig. 1 in Supplementary Materials (SM) shows the calculated structures of Pb(II)/g-C₃N₄, Hg(II)/g-C₃N₄, and Cd(II)/g-C₃N₄.

The dataset of Pb(II)/g-C₃N₄ contains 7000 DFT-based single-point adsorption energies (ΔE). The parallelogram-shaped single layer g-C₃N₄ was fully scanned with respect to the Pb(II) position, as depicted in Fig. 2a. The energy landscape of Pb(II) on the surface of g-C₃N₄ shows that the calculated 7000 ΔE were widely distributed between -0.07 and -4.144 eV with the absolute maximum of 4.144 eV (the black points in Fig. 2a). The randomly placed Pb(II) and g-C₃N₄ have different degrees of adsorption interaction ($\Delta E < 0$), indicating the rationality of the required structural sampling in a real space. Different colors represent different adsorption energies, with the strongest ones locating at the center of a dashed triangle (see discussion in Supplementary Note 1). To reach an accuracy of 0.1 eV^{27,36}, the accurate DNN predictions for ΔE were needed and an appropriate descriptor was selected. To preserve all natural symmetries of the system, a local environment matrix (LEM) was used as a structural descriptor^{37,38}, which is an extensive, continuously differentiable approach and

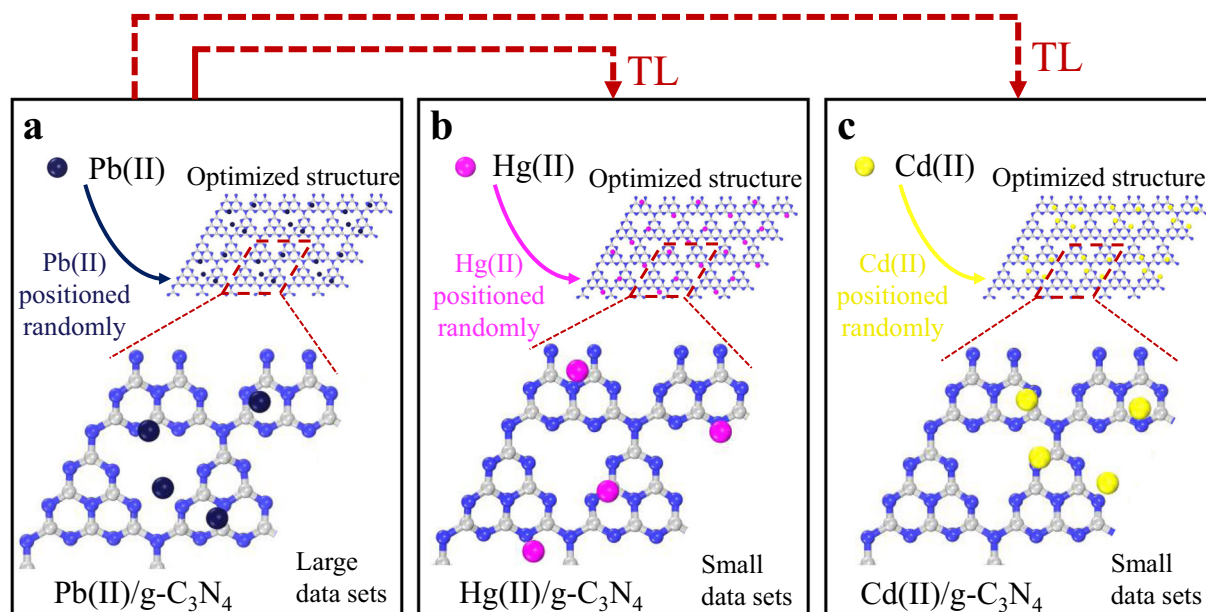


Fig. 1 Schematic model of adsorption process towards HMIs on the surface of g-C₃N₄. (a) Pb(II), (b) Hg(II), and (c) Cd(II). The Pb(II) (black atoms), Hg(II) (pink atoms), and Cd(II) (yellow atoms) were adsorbed randomly at the arbitrary sites of optimized g-C₃N₄. The dataset of Pb(II)/g-C₃N₄ contains 7000 DFT-based adsorption energies (ΔE), with training from scratch, while Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ contain 700 DFT-based adsorption energies (ΔE) respectively, with training based on transfer learning.

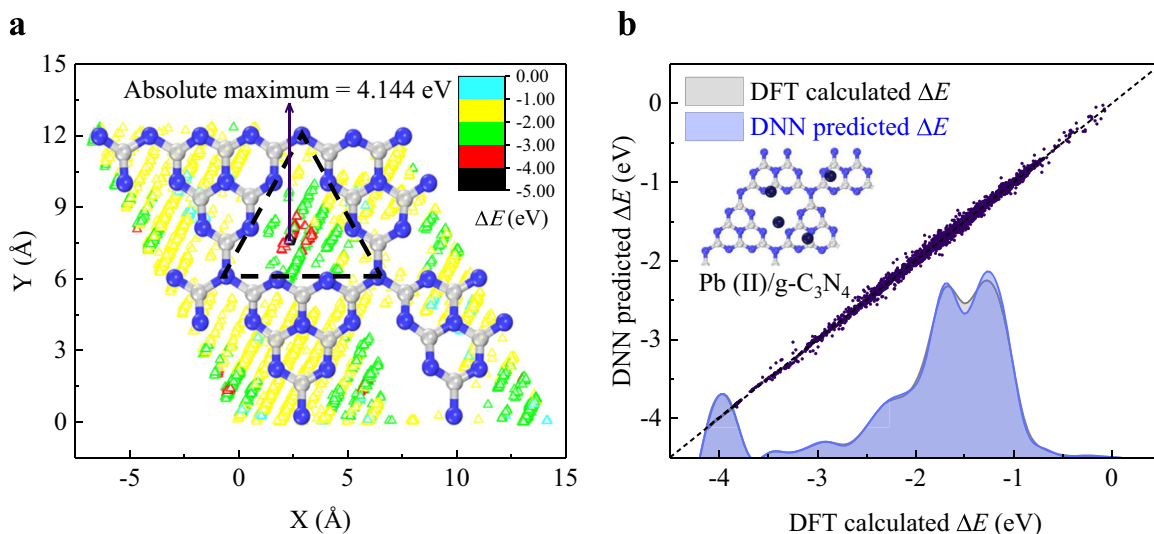


Fig. 2 Dataset of Pb(II)/g-C₃N₄ and ML results. **a** Pb(II) position scan at the arbitrary sites of the surface of a parallelogram-shaped g-C₃N₄, with the corresponding energy landscape calculated by DFT. The blue, yellow, green, red, and black open triangles represent the adsorption energies from −0.07 and −4.144 eV, with the absolute maximum of 4.144 eV. **b** Correlation plot of adsorption energy against DFT and DNN, along with histograms of predicted (blue) and calculated (gray) energy distributions.

Table 1. The RMSEs (eV) and R^2 of Pb(II)/g-C₃N₄, Hg(II)/g-C₃N₄, and Cd(II)/g-C₃N₄, based on DNN and TL methods.

	Datasets		RMSE (eV)				R^2	
	Training	Testing	Training (FS)	Testing (FS)	Training (TL)	Testing (TL)	FS	TL
Pb(II)	6000	1000	0.034	0.051	–	–	0.99	–
Hg(II)	600	100	0.033	0.423	0.014	0.012	0.79	0.99
Cd(II)	600	100	0.099	0.121	0.029	0.043	0.90	0.99

The Pb(II)/g-C₃N₄ structure was predicted with 7000 DFT adsorption energies, while the Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ were predicted with 700 energies, respectively.

linear to the size of the system. Compared to traditional kernels and hand-crafted features, LEM performs well in many systems, such as organic molecules and metal materials²⁸, thus serving as a feature space for DNN input in this study. From Fig. 2a, most of the yellow and green points (with absolute energies below 3 eV) were adsorbed at the edges of the parallelogram-shaped single layer g-C₃N₄, while the red and black points with strong adsorption energies were located in the center from the top view. The position scan of Pb(II) in Fig. 2a shows that Pb(II) is more favorably adsorbed at the center of the g-C₃N₄ rather than at the edges from the top view.

Figure 2b shows the correlation plot of ΔE between DFT and DNN calculations of Pb(II)/g-C₃N₄, including 6000 training points and 1000 testing points. From the dashed line errors, the points are uniformly distributed on both sides of the dashed line around $y = x$ from −0.07 and −4.144 eV. The determination coefficient (R^2) of Pb(II)/g-C₃N₄ model obtained from these scattered points is 0.99 (as shown in Table 1), this indicated that the DNN prediction for energy distribution is in good agreement with the DFT, and the maximum deviation between DFT and DNN is 0.133 eV. Especially, on a single-CPU, it takes only a few milliseconds for DNN to predict an adsorption energy, which is millions of times faster than the DFT calculations. Therefore, our method can not only predict the adsorption ability of g-C₃N₄ towards Pb(II) at the arbitrary site, but also maintains the DFT level of accuracy. In addition, in the training stage, this AI method is ten times faster than the training from scratch (only requires one-tenth of datasets than training from scratch), while millions of times faster than the DFT in the prediction stage.

Adsorption abilities of g-C₃N₄ for Hg(II) and Cd(II)

To evaluate the adsorption ability of g-C₃N₄ towards Pb(II), 7000 adsorption energies were calculated by the DFT method which was a time-consuming but worthwhile process since such sufficient data ensured the prediction accuracy of initial prediction. To maintain the same prediction accuracy as Pb(II)/g-C₃N₄ but shorten the calculation time, we used the TL method to evaluate the adsorption abilities towards Hg(II) and Cd(II). TL enabled the transfer of feature representation learned for a specific predictive modeling task from a large data source set to small target datasets in a similar domain (Fig. 3)^{29–32}, thus it could transfer the DNN prediction of Pb(II)/g-C₃N₄ into similar systems with less data and higher reliability. Compared with the Pb(II)/g-C₃N₄ prediction with 7000 DFT adsorption energies, the adsorption ability predictions for Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ were achieved by the calculations of 700 adsorption energies, respectively. The energy landscapes of Hg(II) and Cd(II) scans at the arbitrary sites of the surface of parallelogram-shaped g-C₃N₄ were plotted in Supplementary Figs. 2, 3, where only 700 adsorption energies (one-tenth of data of Pb(II)/g-C₃N₄) were calculated by the DFT method.

Table 1 shows the predicted root-mean-squared errors (RMSEs) for Pb(II)/g-C₃N₄, Hg(II)/g-C₃N₄, and Cd(II)/g-C₃N₄. The RMSE of 0.1 eV obtained from the prediction model by only a few hundred DFT calculations is a remarkable achievement, which provides a powerful guarantee for the statistical prediction of adsorption capacity of materials to HMIs. As expected, based on the structural descriptor of LEM, we can fleetly predict the adsorption energy of Pb(II) at the arbitrary site with an accuracy of 0.051 eV for the 1000

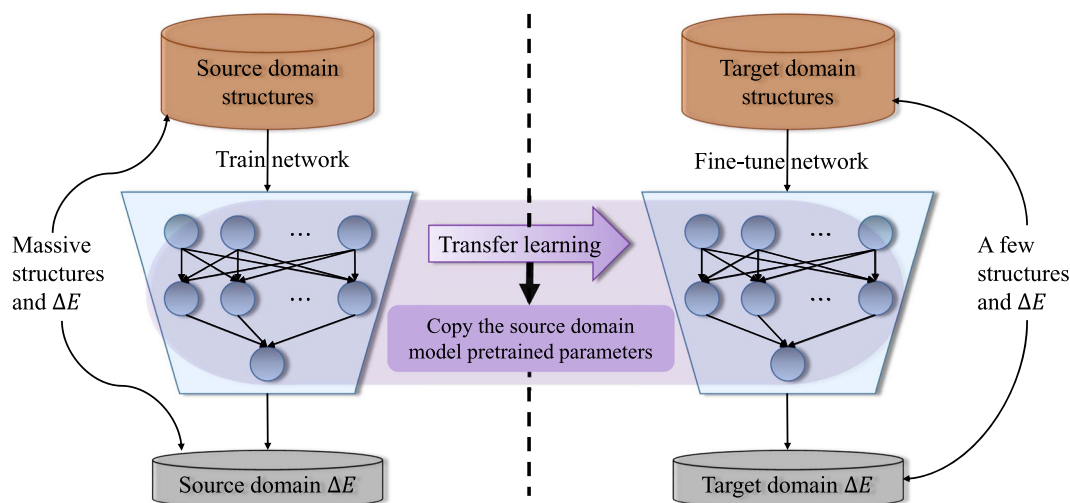


Fig. 3 Schematic diagram of transfer learning. The model of Pb(II)/g-C₃N₄ was chosen as the source domain, with massive structures and ΔE , while the models of Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ were target domains, with a few structures and ΔE . In the training of Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄, the parameters in Pb(II)/g-C₃N₄ model were taken as the starting points for Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄, instead of randomly initializing parameters, and then the parameters were further fine-tuned for training.

testing data, while the testing RMSEs for Hg(II) and Cd(II) are 0.012 eV and 0.043 eV for 100 testing data, respectively.

To clarify the rationality and accuracy of the TL method in processing small datasets, Fig. 4 shows the performance comparison of trained models from scratch (FS) and transfer learning (TL) in each iteration for Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄, based on 700 single-point adsorption energies by DFT calculation. In the FS method, the model parameters were initialized randomly from a uniform distribution and all feature attributes were learned from the input training data. The 700 adsorption energies were randomly divided into 600 training data and 100 testing data. The red and orange curves in Fig. 4a are the training and testing RMSE of Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ based on FS, where the RMSE decreases as the increasing of iterations. However, during the iterations of 200–400 and 1000–2200, the training RMSE based on FS exhibits abnormal decreasing while the testing RMSE increases with the increasing of iterations. This is a typical overfitting effect induced by FS, resulting in large prediction errors for training and test data and poor generalization ability of the model. The same overfitting effect can also be found in the Cd(II)/g-C₃N₄ structure in Fig. 4b. The green and purple curves in Fig. 4a, b show the training and testing RMSEs based on the TL method. Different from the FS method with randomly initialized parameters, the model parameters for Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ coming from TL were initialized based on the well-trained model of Pb(II)/g-C₃N₄ and fine-tuned in the next training. Table 1 lists the training errors, test errors and R^2 of Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ obtained by TL and FS methods, where the testing errors for Hg(II) are 0.423 eV with FS and 0.012 eV with TL, and the ones for Cd(II) are 0.121 eV with FS and 0.043 eV with TL. The prediction RMSEs based on TL are several times smaller than those of the FS, highlighting the advantages of TL. Figure 4c, d shows the correlation plots of adsorption energies between DFT calculations and DNN predictions (based on FS and TL) of Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄, along with the related distributions of ΔE . The R^2 for Hg(II) is 0.79 with FS and 0.99 with TL, and the ones for Cd(II) are 0.90 with FS and 0.99 with TL. The correlation plots of ΔE between DFT (black curves) and TL (pink curve for Hg(II) and yellow curve for Cd(II)) line up over each other, indicating the similar prediction accuracy of TL and DFT, while the blue dots and blue curves based on FS are away from the DFT calculations. Furthermore, in Table S1, for Hg(II), the maximum deviations of DFT and DNN are 1.266 eV with FS and 0.056 eV with TL, and the ones for Cd(II) are

1.515 eV with FS and 0.054 eV with TL. For the training from scratch, the maximum error exceeds 1.0 eV (the relative error is more than 50%), which is likely to result in very inaccurate model predictions.

Table 1 displays the reliability and effectiveness of the TL method, where the prediction errors of Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ based on the TL method and 700 DFT energies are 0.012 eV and 0.043 eV, respectively. Although the 700 data size is small, the prediction errors for Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ are far below than those of Pb(II)/g-C₃N₄ based on FS and 7000 DFT energy (0.051 eV). Therefore, even if the dataset is very small (like the Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ calculations with only about 600 samples for fine-tuning), the proposed TL method can work effectively even if the target domain has a small amount of data, as long as an accurate model is established on the source domain (see Supplementary Note 2). Supplementary Fig. 4 shows the comparison of FS predictions for Pb(II)/g-C₃N₄, Hg(II)/g-C₃N₄, and Cd(II)/g-C₃N₄, with 7000, 700, and 700 adsorption energies, respectively, where the FS performs well for the 7000 energies of Pb(II)/g-C₃N₄, but poorly for the 700 energies of Hg(II)/g-C₃N₄ or Cd(II)/g-C₃N₄. Therefore, the size of training dataset in ML has a significant impact on the model performance^{27,39}, where FS fails to predict the system with small size of dataset, but TL can. More statistical information of adsorption energies are provided in Supplementary Figs. 5–7 and Supplementary Table 1.

Adsorption ability comparison and verification

The above adsorption ability prediction for Pb(II)/g-C₃N₄, Hg(II)/g-C₃N₄, and Cd(II)/g-C₃N₄ was based on 7000, 700, and 700 single-point adsorption energies by DFT calculations, respectively. To compare the adsorption abilities of three HMIs, we filled the datasets of Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ with 6300 adsorption energy points, making them the same sizes as the datasets of Pb(II)/g-C₃N₄, based on the TL prediction rather than DFT calculation. By using the proposed heavy metal ion-transfer learning (HMI-TL) model, such large datasets enable us to obtain the unbiased and reliable statistical results without computational cost. Figure 5 shows the frequency histograms of g-C₃N₄ towards three HMIs, where the blue, pink, and yellow curves represent the energy distributions of Pb(II), Hg(II), and Cd(II) with 7000 adsorption energies, respectively. Table 1 shows the energy distributions of three HMIs adsorbed on the surface of g-C₃N₄ at arbitrary sites,

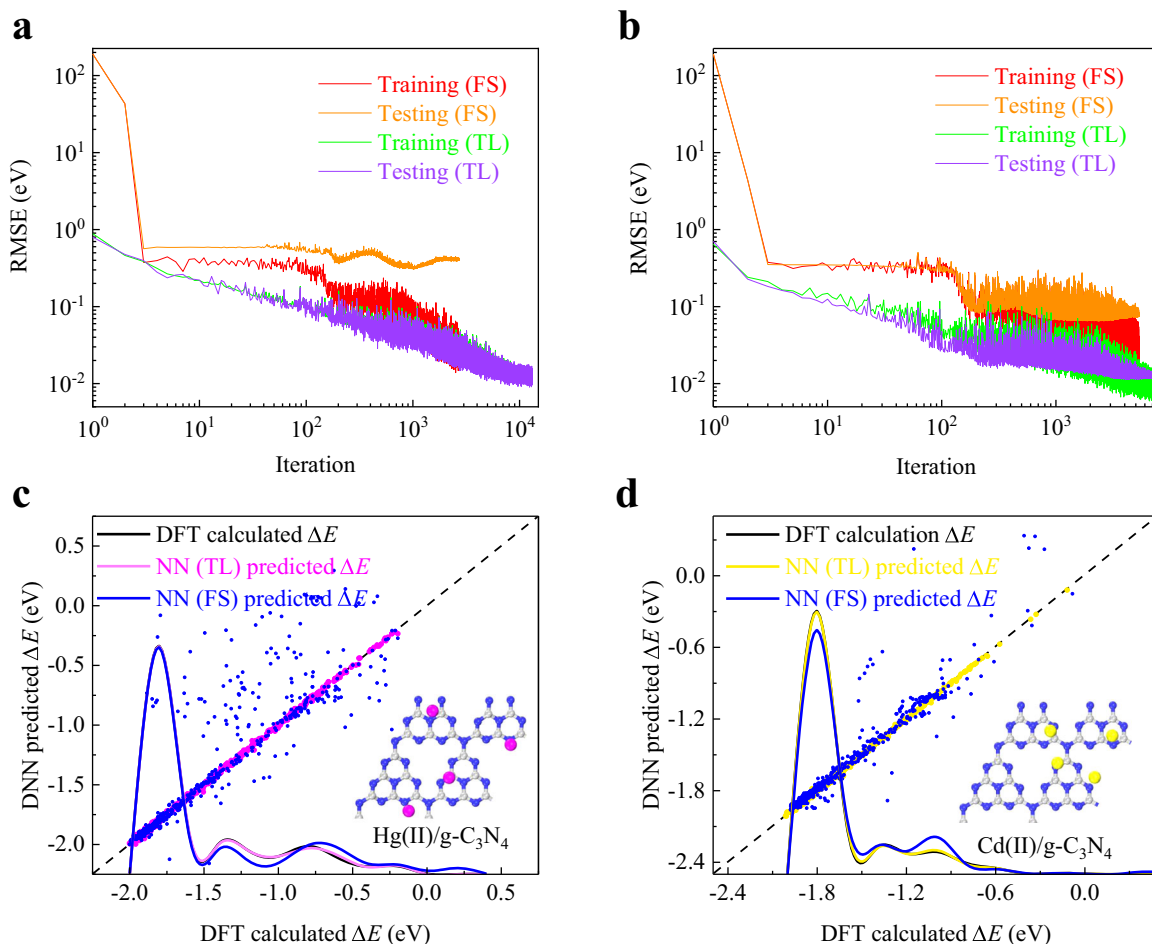


Fig. 4 Performance of the model trained from scratch (FS) and transfer learning (TL). **a** The training and testing RMSE of Hg(II)/g-C₃N₄ as a function of iteration. **b** The training and testing RMSE of Cd(II)/g-C₃N₄ as a function of iteration. **c, d** The correlation plots of ΔE of Hg(II)/g-C₃N₄ (c) and Cd(II)/g-C₃N₄ (d) between DFT calculations and DNN predictions (FS and TL), along with histograms of predicted and calculated energy distributions.

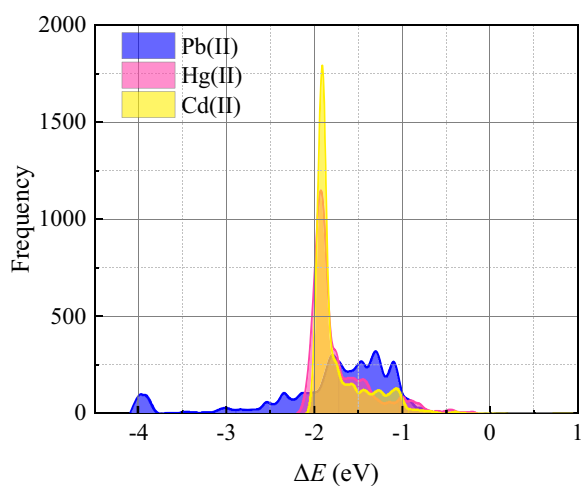


Fig. 5 Overall comparison. Three histograms of g-C₃N₄ towards Pb(II), Hg(II) and Cd(II) with 7000 calculated ΔE by DFT, 700 calculated ΔE by DFT + 6300 predicted ΔE by TL, and 700 calculated ΔE by DFT + 6300 predicted ΔE by TL, respectively.

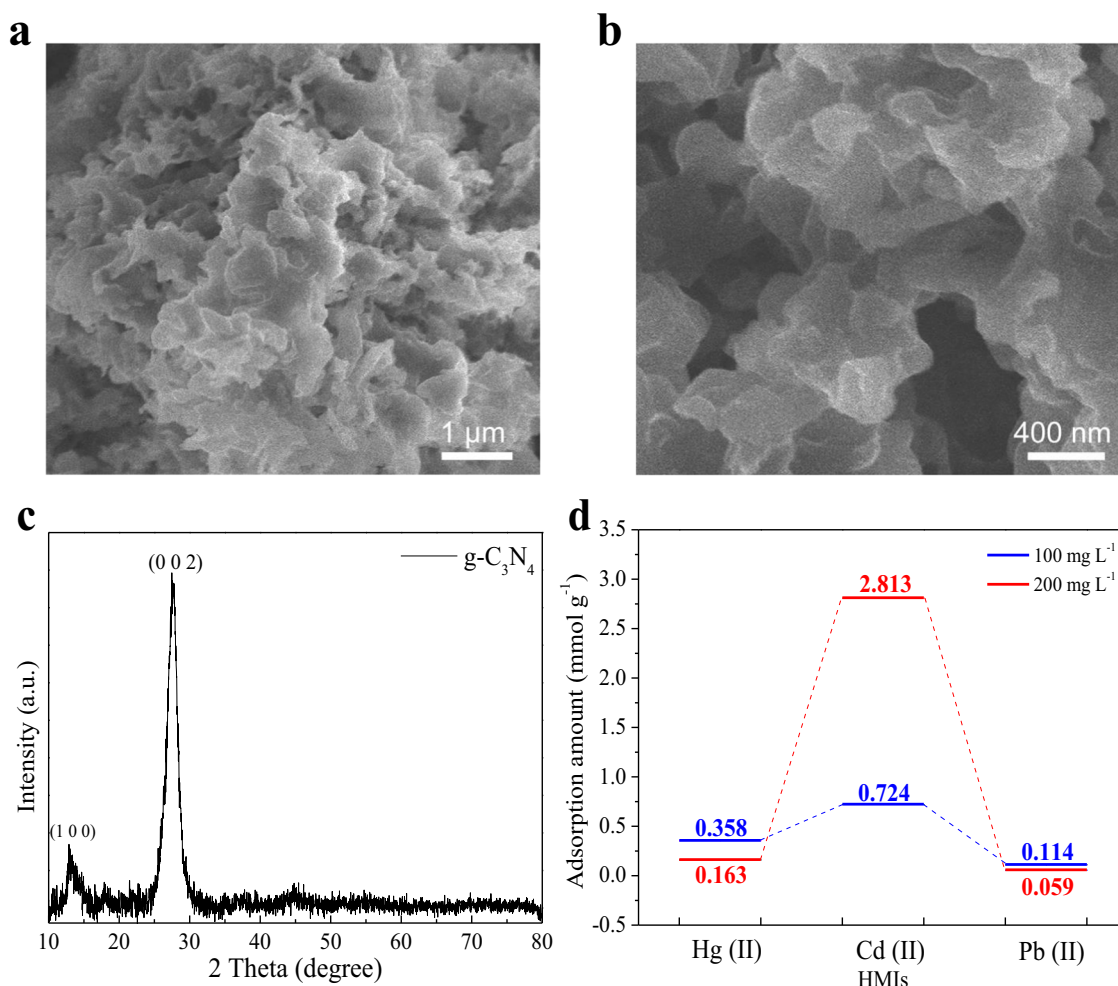
where the adsorption energies of Pb(II) on the surface of g-C₃N₄ distributed between −4.144 eV and −0.07 eV, while the adsorption energies of Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ are distributed between −2.136 eV and −0.139 eV, and −2.048 and −0.051 eV,

respectively. The mean adsorption energies for Pb(II), Hg(II), and Cd(II) are −1.664, −1.695, and −1.707 eV, respectively.

To predict the adsorption ability of a material toward HMIs, the traditional method is to optimize the composite structure to obtain the adsorption energy of HMIs at a fixed position of the material. It is unilateral to evaluate the adsorption ability of a material toward HMIs at a fixed position, only corresponding to one point in Fig. 2a. The proposed study considers the prediction of adsorption energy of HMIs at arbitrary sites of an adsorbent material. From Fig. 5, the predicted and calculated 7000 adsorption energies for each HMI are distributed in different energy ranges, indicating the different adsorption abilities of g-C₃N₄ adsorbent at different positions. In Table 2, the standard deviation shows the energy distribution of three HMIs, among which the Pb(II) with the largest standard deviation has the widest energy distribution, followed by Hg(II) and Cd(II). Such distributions are consistent with the curve trend in Fig. 5. The widely distributed adsorption energy makes it more difficult to evaluate the collective adsorption capacity of a material towards certain HMIs. To evaluate the adsorption abilities of different HMIs on the surface of g-C₃N₄, we calculated the mean of 7000 adsorption energy distributions of each HMI, with the results of −1.664 eV for Pb(II), −1.695 eV for Hg(II), and −1.707 eV for Cd(II), as shown in Table 2. Therefore, based on the HMI-TL model, the adsorption abilities for three ions at any site of the g-C₃N₄ surface are evaluated as Cd(II) > Hg(II) > Pb(II).

Table 2. Adsorption ability comparison of g-C₃N₄ material towards three HMIs.

	Datasets	Minimum (eV)	Maximum (eV)	Standard Deviation (eV)	Mean (eV)
Pb(II)	7000	−4.144	−0.070	0.585	−1.664
Hg(II)	7000	−2.136	−0.139	0.355	−1.695
Cd(II)	7000	−2.048	−0.051	0.318	−1.707

**Fig. 6** Experimental verification. **a**, **b** SEM images of g-C₃N₄ and **c** corresponding XRD pattern. **d** The adsorption ability towards three HMIs in two initial concentrations.

Furthermore, to validate the prediction of the presented HMI-TL model for evaluating the relative adsorption ability on the arbitrary sites of the adsorbent's surface, we have experimentally measured the adsorption capacity of g-C₃N₄ adsorbent towards the three HMIs. The g-C₃N₄ was synthesized by calcining urea at 550 °C. As shown in Fig. 6a, a porous interconnection structure was observed and the high-magnification scanning electron microscopy (SEM) image (Fig. 6b) indicated that the layered g-C₃N₄ presented wrinkled structure, which was beneficial for adsorbing HMIs. Figure 6c shows the X-ray diffraction (XRD) pattern of g-C₃N₄. Two strong peaks at $2\theta = 13.2^\circ$ and 27.6° are observed, which are indexed to the (100) and (002) planes of graphitic nature, respectively^{40,41}.

In our measurements, two initial concentrations (100 and 200 mg L^{−1}) were used for measuring the adsorption ability towards three HMIs, respectively. The adsorption capability of the

g-C₃N₄ was obtained by measuring the concentrations of the solutions before and after adsorption. As plotted in Fig. 6d, the order of adsorption amounts under the same conditions follow the sequence of Cd(II) > Hg(II) > Pb(II) either in 100 or 200 mg L^{−1} solutions, which is consistent with our theoretical prediction, and these results prove that our method is feasible and effective. For the large change of values in Fig. 6d, we thought it may be attributed to the influence of steric hindrance and the interaction between HMIs. Owing to the relatively strong adsorption of g-C₃N₄ for Cd(II), the values change more obviously at different initial concentrations. In this work, we aimed at exploring an ML method to rapidly pick out relatively strong adsorption adsorbents. We expect that it will be significant to further study the practical application of the presented ML method, which would be an important research direction.

DISCUSSION

In this study, we proposed an AI approach to evaluate the adsorption ability of adsorbent toward HMIs at arbitrary sites accurately and quickly, based on the deep neural network and transfer learning. As a case study, we chose a typical g-C₃N₄ as the adsorbent to investigate the adsorption abilities toward three representative HMIs (Pb(II), Hg(II), Cd(II)). The Pb(II) on g-C₃N₄ was evaluated by 7000 single-point adsorption energies with DFT calculations, while the Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ were calculated by only 700 DFT energies with the TL method. The predicted adsorption abilities for the three HMIs were Cd(II) > Hg(II) > Pb(II), corresponding to the predicted RMSE of 0.043, 0.012, and 0.051 eV, respectively. Such RMSEs, all less than 0.1 eV with only a few hundred DFT calculations, ensured the prediction accuracy and were considered as a remarkable feat. Furthermore, the predicted results are also confirmed by experimentally measuring the adsorption efficiency of g-C₃N₄ adsorbent towards Cd(II), Hg(II), and Pb(II).

While significant research progress has been achieved by finding and designing adsorbents to deal with the water pollution, the prediction of adsorption ability of adsorbents to HMIs is still a challenge. First, the experimental prediction of adsorption capacity of adsorbents on HMIs is a complex process, involving multiple steps such as material design, synthesis, and measurement. It is really difficult to determine the adsorption capacity of adsorbents at arbitrary sites. Second, the first-principles calculations can quantitatively determine the adsorption sites and ability of different adsorbents to HMIs, but it is time-consuming. In our study, to obtain the adsorption ability of any site, the presented AI approach can reach the same prediction accuracy as the first-principles calculation, but only requires one-tenth of datasets than training from scratch, which means it is ten times faster than the training from scratch in the training stage and millions of times faster than the DFT in prediction stage. The present study shows that the HMI-TL model can accurately and rapidly evaluate the adsorption ability of the adsorbent towards HMIs and determine the adsorption position at arbitrary sites without involving the experimental process. HMT-TL model provides a convincing and powerful pre-experimental guidance for removing of certain HMIs, which is of great significance to design adsorption materials.

To sum up, this work has demonstrated the feasibility of transfer learning in evaluating the adsorption capacity of adsorbent materials for HMIs, based on a small amount of data. When the source field used for transferring learning is similar to the target field, we believe that the proposed HMI-TL model can effectively transfer knowledge from the source dataset to the target dataset. In addition, the AI approach proposed in this work can help solve HMIs and organic contamination in aqueous solutions, which can be used to screen more robust materials when designing and discovering adsorbents. Considering that the prediction of adsorption processes can be widely used in many fields such as catalysis and batteries^{42,43}, the proposed model provides an opportunity to solve adsorption problems by combining AI, materials, and environmental science.

METHODS

First-principles calculations

The DFT calculations were conducted using the Vienna ab initio Simulation Package (VASP)⁴⁴. The projected augmented wave (PAW) method^{45,46} was applied to describe ion-electron interactions along with the Perdew–Burke–Ernzerhof (PBE) exchange–correlation function within generalized gradient approximation (GGA). The Hg, Cd, Pb, and single layer g-C₃N₄ were optimized in advance. During the adsorption calculation, a cutoff energy of 500 eV was performed with a Monkhorst-Pack of 3 × 3 × 1 k-point grids, and the convergence criteria were set to 1 × 10^{−6} eV atom^{−1} for energy and 0.01 eV Å^{−1} for force, respectively. A vacuum distance of 15 Å was added in the g-C₃N₄ slab to avoid periodic interactions. To

accurately describe the interaction between HMIs and g-C₃N₄ substrate, the DFT-D3 method⁴⁷ was employed, which considers the van der Waals interaction. The adsorption energy can be described as

$$\Delta E = E_{\text{sub} + \text{met}} - E_{\text{sub}} - E_{\text{met}} \quad (1)$$

where $E_{\text{sub} + \text{met}}$, E_{sub} and E_{met} were the energy of HMIs adsorbed on the surface of the substrate, the energy substrate of g-C₃N₄ and the energy of HMIs (Cd, Hg, Pb), respectively.

Datasets

The three datasets for Pb(II)/g-C₃N₄, Hg(II)/g-C₃N₄, and Cd(II)/g-C₃N₄ contained 7000, 700, and 700 single adsorption energies calculated by DFT, respectively. The HMIs were randomly scanned on the parallelogram-shaped g-C₃N₄. To explore the adsorption active sites and make the adsorption energy negative, Pb(II) were randomly scanned at a distance of 100–300 pm from the surface, where the seed value of rand function was changed by using the system time and the different random number sequences were generated by C++ program. Furthermore, the datasets of Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄ were produced in a similar fashion, where the Hg(II) and Cd(II) were randomly scanned at a distance of 200–400 pm from the surface.

Structural descriptors

Structural descriptors are the input vectors of NN, satisfying the translational, rotational, and permutational invariance. In this study, we used a LEM as the structural descriptor³⁷, which is an extensible approach and has powerful functions. For a n atoms system, the Cartesian coordinates are $\{\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n\}$, where $\mathbf{R}_i = \{x_i, y_i, z_i\}$, \mathbf{R}_{ij} is the vector of $\mathbf{R}_i - \mathbf{R}_j$. We calculated the entire radial and angular features of atom i and neighbor atom j base on the equation:

$$G_{ij} = \left\{ \frac{1}{V_{ij}}, \frac{x'_{ij}}{V_{ij}^2}, \frac{y'_{ij}}{V_{ij}^2}, \frac{z'_{ij}}{V_{ij}^2} \right\} \quad (2)$$

where $x'_{ij}, y'_{ij}, z'_{ij}$ could be obtained by x_{ij}, y_{ij}, z_{ij} through rotation matrix \mathfrak{R} , which could be expressed as

$$V_{ij} = \{x'_{ij}, y'_{ij}, z'_{ij}\} = \{x_{ij}, y_{ij}, z_{ij}\} \mathfrak{R}(\mathbf{R}_{ia(i)}, \mathbf{R}_{ib(i)}) \quad (3)$$

The rotation matrix \mathfrak{R} were defined by the two closest atoms (atom ia and ib), independently of their chemical elements, and atom i :

$$\mathfrak{R}(\mathbf{R}_{ia}, \mathbf{R}_{ib}) = \begin{pmatrix} e(\mathbf{R}_{ia}) \\ e[\mathbf{R}_{ib} - (\mathbf{R}_{ia} \cdot \mathbf{R}_{ib})\mathbf{R}_{ia}] \\ e(\mathbf{R}_{ia} \times \mathbf{R}_{ib}) \end{pmatrix}^T \quad (4)$$

where $e(x) = x/\|x\|$. The rotation matrix \mathfrak{R} could also be named as local frame of atom i . Therefore, different atoms had different rotation matrices. We set 10.0 Å as the cutoff radius for neighbor searching and 8.8 Å as where the smoothing started.

Training model

The Deep Potential-Smooth Edition (DeepPot-SE) model implemented by Python/C++ and TensorFlow framework⁴⁸ was used in this study. DeepPot-SE, an end-to-end DNN-based PES model, which is able to efficiently represent the PES of a wide variety of systems with the accuracy of ab initio quantum mechanics. It is extensive and continuously differentiable, scales linearly with system size, and preserves all the natural symmetries of the system. In the model, the three hidden layers each with 20 nodes were fully connected, which was determined by the DNN to predict the adsorption energies via structural descriptors. A batch size of 64 with Adam optimizer was used to improve the training speed while strengthening the optimization⁴⁹. During the training, the error of the model was tested and displayed every 100 iterations. We used the initial learning rate of 0.002 for the model that was trained from scratch and 0.001 for the model based on TL, given that hyper parameters were fine-tuned during the TL training.

In this work, we used the parameter-based TL. The source domain and the target domain share model parameters, that is, the model trained by a large amount of data in the source domain is applied to the target domain for prediction. The parameter-based TL method is more straightforward and has the advantage of making full use of the similarity between models. In this work, Pb(II)/g-C₃N₄ is a well-trained model based on large datasets. Before training, the parameters of the Pb(II)/g-C₃N₄ model are randomly

initialized. In the training of Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄, the parameters in Pb(II)/g-C₃N₄ model are taken as the starting points for Hg(II)/g-C₃N₄ and Cd(II)/g-C₃N₄, instead of randomly initializing parameters, and then the parameters are further fine-tuned for training.

Experimental validation

The g-C₃N₄ adsorbent was prepared by heating 10 g of urea in a crucible for 3 h under a heating speed of 5 °C min⁻¹. The HMLs solutions with initial concentrations of 100 and 200 mg L⁻¹ were prepared by using Cd(NO₃)₂, Hg(NO₃)₂, Pb(NO₃)₂ as sources, respectively. To reach the adsorption equilibrium, these mixtures were shaken for 24 h and the suspensions were centrifuged. The residual concentrations of HMLs were measured by an inductive coupled plasma (ICP) atomic emission spectrometer (Optima 7300 DV, USA). The adsorption amount of HMLs in g-C₃N₄ could be obtained by the formula (mmol g⁻¹):

$$q_e = \frac{C_0 - C_e}{Mm} \times V \quad (5)$$

where C_0 is the initial concentration of HMLs, C_e is the residual concentration of HMLs, M is the relative atomic mass of Cd, Hg, and Pb, V is the volume of adsorption solution, and m is the mass of g-C₃N₄ adsorbent.

DATA AVAILABILITY

All datasets generated in the current study are available from the corresponding author upon reasonable request.

CODE AVAILABILITY

The codes of generating the structures in this study are available from the corresponding author upon reasonable request. The codes of DeepPot-SE model used in this study is available at <https://github.com/deepmodeling/deepmd-kit>.

Received: 26 June 2020; Accepted: 6 January 2021;

Published online: 29 January 2021

REFERENCES

- Lu, S. et al. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. Commun.* **9**, 3405 (2018).
- Yuan, R. et al. Accelerated discovery of large electrostrains in BaTiO₃-based piezoelectrics using active learning. *Adv. Mater.* **30**, 1702884 (2018).
- Zhang, Y. et al. Unsupervised discovery of solid-state lithium ion conductors. *Nat. Commun.* **10**, 5260 (2019).
- Xu, J. et al. Remediation of heavy metal contaminated soil by asymmetrical alternating current electrochemistry. *Nat. Commun.* **10**, 2440 (2019).
- Bolisetty, S., Peydayesh, M. & Mezzenga, R. Sustainable technologies for water purification from heavy metals: review and analysis. *Chem. Soc. Rev.* **48**, 463–487 (2019).
- Jia, L. et al. Interactions of high-rate nitrate reduction and heavy metal mitigation in iron-carbon-based constructed wetlands for purifying contaminated ground-water. *Water Res.* **169**, 115285 (2020).
- Zheng, S., Wang, Q., Yuan, Y. & Sun, W. Human health risk assessment of heavy metals in soil and food crops in the Pearl River Delta urban agglomeration of China. *Food Chem.* **316**, 126213 (2020).
- Hu, C. et al. Carbon-based metal-free catalysts for energy storage and environmental remediation. *Adv. Mater.* **31**, 1806128 (2019).
- Liu, C. et al. Direct/Alternating current electrochemical method for removing and recovering heavy metal from water using graphene oxide electrode. *ACS Nano* **13**, 6431–6437 (2019).
- Jiang, Y., Liu, C. & Huang, A. EDTA-functionalized covalent organic framework for the removal of heavy-metal ions. *ACS Appl. Mater. Interfaces* **11**, 32186–32191 (2019).
- Zhou, L. et al. Effective removing of hexavalent chromium from wasted water by triboelectric nanogenerator driven self-powered electrochemical system—why pulsed DC is better than continuous DC? *Nano Energy* **64**, 103915 (2019).
- Sun, G. L., Reynolds, Erin, E. & Belcher, A. M. Using yeast to sustainably remediate and extract heavy metals from waste waters. *Nat. Sustain.* **3**, 303–311 (2020).
- Guo, Y. et al. Biomass-derived hybrid hydrogel evaporators for cost-effective solar water purification. *Adv. Mater.* **32**, 1907061 (2020).
- Zhou, X. et al. Steering surface reaction at specific sites with self-assembly strategy. *ACS Nano* **11**, 9397–9404 (2017).
- Briggs, N. M. et al. Identification of active sites on supported metal catalysts with carbon nanotube hydrogen highways. *Nat. Commun.* **9**, 3827 (2018).
- Sellaoui, L. et al. Understanding the adsorption of Pb²⁺, Hg²⁺ and Zn²⁺ from aqueous solution on a lignocellulosic biomass char using advanced statistical physics models and density functional theory simulations. *Chem. Eng. J.* **365**, 305–316 (2019).
- Wang, R. et al. K_x[Bi_{4-x}Mn_xS₆], design of a highly selective ion exchange material and direct gap 2D semiconductor. *J. Am. Chem. Soc.* **141**, 16903–16914 (2019).
- Huang, Q.-S. et al. Highly-efficient Pb²⁺ removal from water by novel K₂W₄O₁₃ nanowires: Performance, mechanisms and DFT calculation. *Chem. Eng. J.* **381**, 122632 (2020).
- Yuan, Y. et al. Frontispiece: a bio-inspired nano-pocket spatial structure for targeting uranyl capture. *Angew. Chem. Int. Ed.* **59**, 4262–4268 (2020).
- Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **8**, 13890 (2017).
- Ziletti, A., Kumar, D., Scheffler, M. & Ghiringhelli, L. M. Insightful classification of crystal structures using deep learning. *Nat. Commun.* **9**, 2775 (2018).
- Winter, R., Montanari, F., Noé, F. & Clevert, D.-A. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chem. Sci.* **10**, 1692–1701 (2019).
- Chen, C. et al. A critical review of machine learning of energy materials. *Adv. Energy Mater.* **10**, 1903242 (2020).
- Cichos, F., Gustavsson, K., Mehlig, B. & Volpe, G. Machine learning for active matter. *Nat. Mach. Intell.* **2**, 94–103 (2020).
- Ng, M.-F., Zhao, J., Yan, Q., Conduit, G. J. & Seh, Z. W. Predicting the state of charge and health of batteries using data-driven machine learning. *Nat. Mach. Intell.* **2**, 161–170 (2020).
- Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
- Jäger, M. O. J., Morooka, E. V., FedericiCanova, F., Himanen, L. & Foster, A. S. Machine learning hydrogen adsorption on nanoclusters through structural descriptors. *npj Comput. Mater.* **4**, 37 (2018).
- Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **5**, 83 (2019).
- Singh, J., Hanson, J., Paliwal, K. & Zhou, Y. RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat. Commun.* **10**, 5407 (2019).
- Jha, D. et al. Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. *Nat. Commun.* **10**, 5316 (2019).
- Smith, J. S. et al. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **10**, 2903 (2019).
- Zhou, Z., Ye, C., Wang, J. & Zhang, N. R. Surface protein imputation from single cell transcriptomes by deep neural networks. *Nat. Commun.* **11**, 651 (2020).
- Zhao, G., Li, J., Ren, X., Chen, C. & Wang, X. Few-layered graphene oxide nanosheets as superior sorbents for heavy metal ion pollution management. *Environ. Sci. Technol.* **45**, 10454–10462 (2011).
- Perreault, F., Fonseca de Faria, A. & Elimelech, M. Environmental applications of graphene-based nanomaterials. *Chem. Soc. Rev.* **44**, 5861–5896 (2015).
- Kumar, P. et al. C₃N₅: a low bandgap semiconductor containing an azo-linked carbon nitride framework for photocatalytic, photovoltaic and adsorbent applications. *J. Am. Chem. Soc.* **141**, 5415–5436 (2019).
- Ma, X., Li, Z., Achenie, L. E. K. & Xin, H. Machine-Learning-augmented chemisorption model for CO₂ electroreduction catalyst screening. *J. Phys. Chem. Lett.* **6**, 3528–3533 (2015).
- Zhang, L., Han, J., Wang, H., Car, R. & E, W. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **120**, 143001 (2018).
- Jackson, N. E. et al. Electronic structure at coarse-grained resolutions from supervised machine learning. *Sci. Adv.* **5**, eaav1190 (2019).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Hou, Y. et al. N-doped graphene/porous g-C₃N₄ nanosheets supported layered-MoS₂ hybrid as robust anode materials for lithium-ion batteries. *Nano Energy* **8**, 157–164 (2014).
- Majumder, S., Shao, M., Deng, Y. & Chen, G. Ultrathin sheets of MoS₂/g-C₃N₄ composite as a good hosting material of sulfur for lithium-sulfur batteries. *J. Power Sources* **431**, 93–104 (2019).
- Deng, D. R. et al. Enhanced adsorptions to polysulfides on graphene-supported BN nanosheets with excellent Li-S battery performance in a wide temperature range. *ACS Nano* **12**, 11120–11129 (2018).

43. Yang, T. et al. High-throughput identification of exfoliable two-dimensional materials with active basal planes for hydrogen evolution. *ACS Energy Lett.* **5**, 2313–2321 (2020).
44. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
45. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
46. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
47. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154104 (2010).
48. Zhang, L. et al. End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. Preprint at <https://arxiv.org/abs/1805.09003> (2018).
49. Kingma, D. & Ba, J. Adam: a method for stochastic optimization. Preprint at <https://arxiv.org/abs/1412.6980> (2017).

ACKNOWLEDGEMENTS

The authors are grateful for the financial support provided by the National Natural Science Foundation of China (No. 21901157), the SJTU Global Strategic Partnership Fund (2020 SJTU-HUJI), the Science and Technology Major Project of Anhui Province (No. 18030901093), Key Research and Development Program of Wuhu (No. 2019YF07), and the Foundation of Anhui Laboratory of Molecule-Based Materials (No. FZJ19014).

AUTHOR CONTRIBUTIONS

Z.W. and H.Z. contribute equally to this work. Z.W., J.R., and J.L. (Jinjin Li) performed the modeling. H.Z., X.L., T.H., and J.L. (Jinyun Liu) conceived of and conducted the experiments. Z.W. and J.L. (Jinjin Li) performed data management. Z.W., H.Z., J.R., T.H., J.L. (Jinyun Liu) and J.L. (Jinjin Li) interpreted the results. J.L. (Jinyun Liu) and J.L. (Jinjin Li) supervised the work. All authors edited and reviewed the manuscript.

COMPETING INTERESTS

The authors declare no financial and/or non-financial competing interests. J.J.L., Z.L. W., H.K.Z., J.H.R., and J.Y.L. have filed a patent related to this work. (Chinese application no. 2020103902417, dated 08 May 2020).

ADDITIONAL INFORMATION

The online version contains supplementary material available at <https://doi.org/10.1038/s41524-021-00494-9>.

Correspondence and requests for materials should be addressed to J.L. or J.L.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021