ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Structure−Energy Relationship Prediction of the HZSM-5 Zeolite with Different Acid Site Distributions by the Neural Network Model

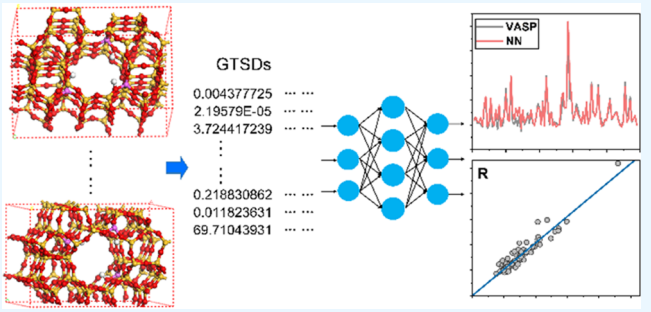
Yang Yang, Wenming Zhang, Shengbin Chen, Xiaogang Wang, Yuangu Xia, Ji Liu, Bin Hu,* Qiang Lu, and Bing Zhang*

Cite This: *ACS Omega* 2024, 9, 3392−3400

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Zeolites are a very important family of catalysts. The catalytic activity of zeolites depends on the distribution of acid sites, which has been extensively studied. However, the relationship between the acid site distribution and catalytic efficiency remains unestablished. An onerous computational burden can be imposed when static calculations are applied to analyze the relationship between a catalyst structure and its energy. To resolve this issue, the current work uses neural network (NN) models to evaluate the relationship. By taking the typical HZSM-5 zeolite as an example, we applied the provided atomic coordinates to predict the energy. The network performances of the artificial neural network (ANN) and high-dimensional neural network (HDNN) are compared using the trained results from a dataset containing the identical number of acid sites. Furthermore, the importance of the feature is examined with the aid of a random forest model to identify the pivotal variables influencing the energy. In addition, the HDNN is employed to forecast the energy of an HZSM-5 system with varying numbers of acid sites. This study emphasizes that the energy of zeolites can be rapidly and accurately predicted through the NN, which can assist our understanding of the relationship between the structure and properties, thereby providing more accurate and efficient methods for the application of zeolite materials.



1. INTRODUCTION

Zeolites are usually crystalline aluminosilicates with uniform channel architecture, which are formed by corner-sharing connection of aluminum−oxygen and silicon−oxygen tetrahedral TO_4 via bridging oxygen.[1,2] They possess a large specific surface area, good hydrothermal stability, and site-specific characteristics of adjustable surface Brønsted acids and Lewis acids,[3] making them versatile catalysts and adsorbents across a broad range of applications, including biomass upgrading,[4] oil refining, thermal energy storage,[5] air pollution remediation,[6] water purification,[7−9] *etc.*

When Al is replaced by Si in the framework of the zeolite, it causes the whole system to carry negative charges and further produce Brønsted acids by means of introducing protons to keep the framework charge neutral,[10,11] thus endowing it with catalytic activity. There are multiple types of acid sites in zeolite, namely, 10-membered ring (10-MR) channels and channel intersections, leading to distinct thermodynamic properties.[12] Hence, the acid site dispersion of zeolites plays an essential role in their catalytic efficiency. The spread of acid sites is closely related to the location of Al.[13] However, due to the limitations of current characterization techniques,[14] the position of Al uncertainty leads to the difficulty in establishing the relationship between the structure and effectiveness of zeolite catalysts.[15] The solution to this challenging issue lies in whether Al at diverse spots of the zeolite has a stable structure with a relatively low energy. Although energy calculations have been undertaken for zeolites at different acid sites, which are generally conducted by means of quantum mechanics-based calculations, there remain certain problems. First, the unit cell of the zeolite is a large system that comprises 288 heavy atoms, imposing a significant computational burden and onerous hardware requirements. Second, there are multiple combinations of acid sites for the zeolite: for the 10-MR of the zeolite, a single acid site shows 12 types of combinations and dual sites demonstrate 194 types of combinations, while the combinations of four sites involve more than 30,000 types. One-by-one evaluations of these combinations are computationally intensive and time-consuming. Nevertheless, establishing the structure−activity relationship between the acid site distributions and structural energy of zeolites can achieve quick and reliable

Received: September 5, 2023
Revised: November 30, 2023
Accepted: December 13, 2023
Published: January 8, 2024

estimation of stability of large zeolite systems at distinct sites.[16,17] This has theoretical and guiding significance in exploring the relationship between the structures of zeolites and their catalytic activity.

Machine learning, as a data-driven technique, has risen to prominence among physicists, chemists, and material scientists.[18] Based on existing databases, machine learning is deployed to estimate various properties of materials. The precision of its projection is equivalent to that of first-principles calculations, but machine learning is expedited by several orders of magnitude in contrast to the latter.[17,19−22] Daeyaert et al.[23] leveraged neural network (NN) models to calculate the stabilization energy of organic structure-directing agents (OSDAs) for the β-zeolite. A data sample with 4781 OSDAs is harnessed to train and validate the NN model for modeling their stabilization energy. A computational speed improvement of 348 times was achieved for stable energy results generated by the NN compared to those calculated by molecular dynamics based on the MMFF force field under the same dataset. Juybar et al.[24] employed an NN to predict the zeolite acidity changes loaded with multiple metals. The acidity changes predicted through the NN were compared to those determined by Fourier transform infrared spectroscopy (FTIR). The outcome indicated that NN predictions align with the results obtained from FTIR spectroscopy. It can be inferred that NNs can rapidly and precisely predict various properties of zeolites. Previous research revealed that the distribution of acid sites is usually a principal reason for zeolite properties including acidity.[25] Rapidly determining the dispersion of acid sites is decisive when understanding various properties of zeolite.

In the zeolite-catalyzed process, replacing atoms with dissimilar elements at the corresponding coordinates or adding an atom in the system due to uncertainty of acid site distribution and acid site density[16] can change the size of the zeolite system.[17] Despite that the ANN estimates the total energy of the whole zeolite system, it can only be used to predict the systems comprising equal numbers of atoms.[26,27] This issue is solved by the HDNN, which leverages the total energy to represent the atomic energy of all atom contributions dependent on chemical environments. The energy contribution of each atom is evaluated by means of a single NN,[17,28] which avoids the drawback necessitating the onerous calculation of the whole system.[26] Therefore, the HDNN is adaptable to the anticipation of systems with any arbitrary number of atoms. Furthermore, to investigate the primary factors that shape system energy, the feature analysis method of random forest is implemented to attain crucial features that influence the energy.[29]

Therefore, this study is based on the given atomic coordinates to predict the energy of HZSM-5 structures with varying distributions of acid sites containing the same number of atoms. Through comparison of the NN prediction results with theoretical calculations, the network performance of the ANN and HDNN is evaluated. The key features that affect the energy are screened for this dataset. Additionally, the HDNN is adopted to project the formation energy of the zeolite with a varying number of atoms. This research contributes to a deeper understanding of the more complex and diverse characteristics of zeolites.

2. METHODOLOGY

2.1. Data Sources. Si or A1 atoms connecting oxygen in the Zeolite Socony Mobil (ZSM)-5 zeolite can be called T tetrahedral atoms, with 12 types included.[30] The ZSM-5 crystal structure[31] is illustrated in Figure 1. The replacement of Si with one, two, and three Al atoms in the zeolites was conducted to ensure diversity of the data, leading to the construction of HZSM-5 model structures with one, two, and three Brønsted acid sites distributed throughout, respectively, resulting in a total of 2501 structures. The energy of each structure was calculated one by one, providing an output data pool for NN computation. A dataset of 2501 data groups was created, each containing atomic coordinates and corresponding energies calculated individually for each structure.

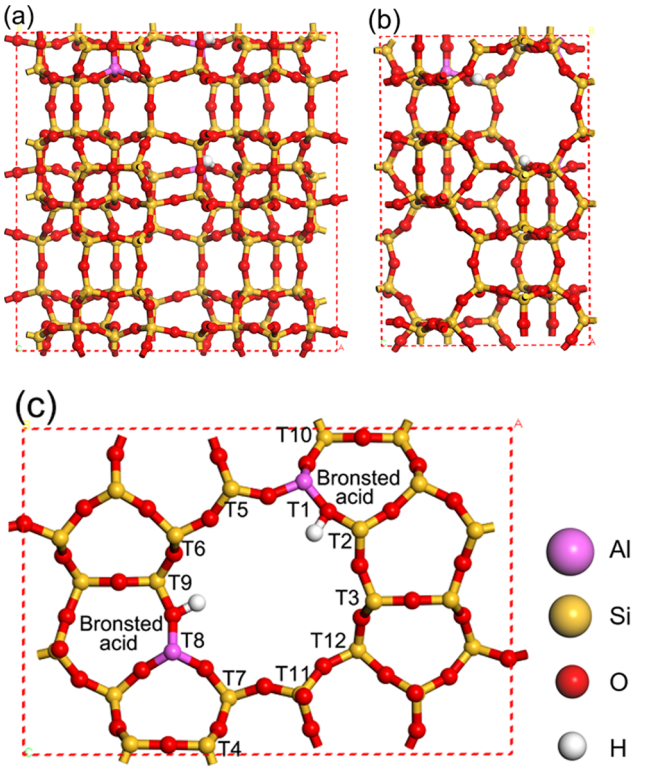


Figure 1. Front view (a), side view (b), vertical view, and distribution of the T sites (c) of the structure of ZSM-5.

All the HZSM-5 structures utilize the same framework, while only the number and location of acid sites were altered. The system energy was calculated using the Vienna *Ab initio* Simulation Package (VASP).[32] Calculations were performed using the Perdew−Burke−Ernzerhof (PBE) exchange-correlation functional.[33] The projected augmented wave (PAW) plane-

Table 1. *R*, RMSE, and Training Time Cost of Network Architectures

network architecture	R_{train}	R_{test}	RMSE (Ha)	training time cost (s)
[10,10]	97.90%	93.62%	0.026	81
[15,15]	99.23%	95.0%	0.020	100
[50,50]	97.79%	83.57%	0.032	220
[100,100]	96.81%	66.72%	0.573	332
[200,200]	93.08%	48.58%	1.672	574
[10,10,10]	98.71%	95.01%	0.021	81
[15,15,15]	97.55%	93.65%	0.032	115
[50,50,50]	96.35%	90.87%	0.039	244
[100,100,100]	97.68%	66.30%	0.547	369
[200,200,200]	95.29%	46.64%	1.190	627

https://doi.org/10.1021/acsomega.3c06689
ACS Omega 2024, 9, 3392−3400

Table 2. Training Results for Different Activation Functions

activation function	R_{train}	R_{test}	RMSE (Ha)	training time cost (s)
[tansig, tansig]	99.23%	95.0%	0.020	100
[logsig, logsig]	97.44%	89.95%	0.033	118
[tansig, logsig]	97.96%	93.16%	0.024	127
[logsig, tansig]	98.21%	94.13%	0.019	127

wave basis was used to describe the atomic properties.[34] The plane-wave cutoff energy was set to 520 eV. The γ-method was adopted to establish a 2 × 2 × 2 k-point grid to ensure convergence of the system energy and force to standard values. The nonlocal van der Waals (vdW) interactions were computed with Grimmer's DFT-D3 (BJ) correction.[35]

2.2. Selection of Parameters. All ANN models were realized utilizing the MATLAB package,[36] while HDNN training was implemented in the RuNNer code.[37] The random forest model was achieved using scikit-learn.[38] Gaussian-type structure descriptors (GTSDs)[39] were taken as the input, and the energy was taken as the output. The GTSDs were used to describe the local chemical environment around each atom through an atomic-centered symmetry function. The distance between the central and neighboring atoms is determined implementing a cutoff radius, which is chosen to be large enough

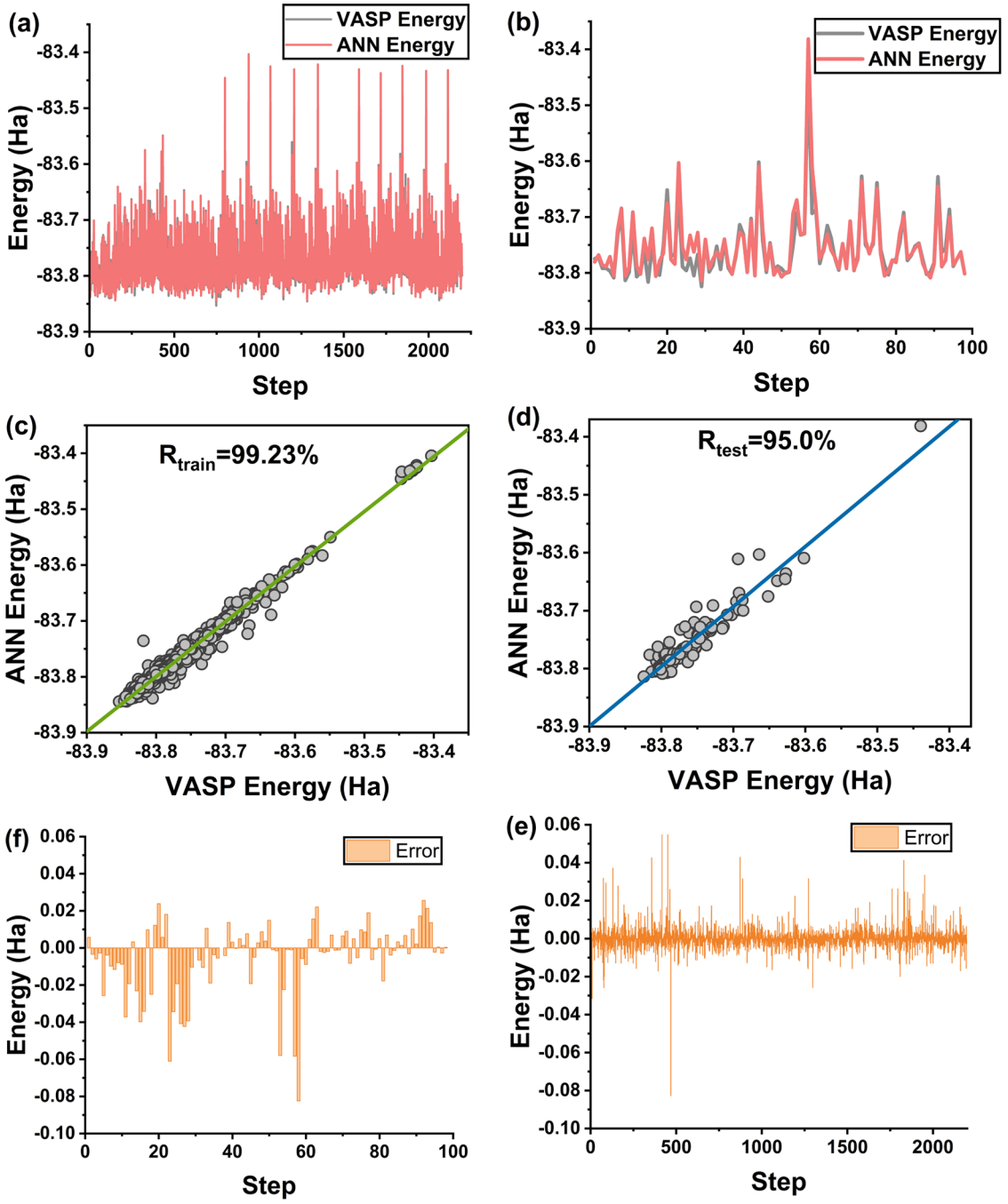


Figure 2. Energy curves of the training set (a) and testing set (b), the correlation between the training set (c) and testing set (d), and the error distribution for the training set (e) and test set (f) of the [15, 15] ANN model.

https://doi.org/10.1021/acsomega.3c06689
ACS Omega 2024, 9, 3392−3400

to determine the maximum possible interactions between them. In this work, a cutoff radius of 18 Bohr is employed to examine two-body and three-body interactions with all other elements, including the central atom, with a total of 104 functions used to describe each atom.

The NNs are trained for energy prediction by utilizing GTSDs based on atomic coordinates as input data. Due to the limitations of the traditional ANN, it is only capable of training a system with uniform numbers of atoms. For HZSM-5, there are only 12 types of single-site structures and 194 types of dual-site structures. However, the size of the dataset composed of these structures is significantly insufficient. Therefore, the HZSM-5 system with the largest number of structures (2295 sets of structural energy data), including only three acid sites, was selected for training. In contrast, the HDNN can be used to predict the energy of different system sizes; hence, HZSM-5 structure−energy data containing 1−3 acid sites (a total of 2501 sets) were used for training the HDNN model. The proportions of the training and test sets of two types of NN models are both 96:4.

Aiming at the screening of variables for the models, at first, the numbers of hidden layers, nodes, and activation functions are tested. The two and three hidden layers are tested, with 10, 15, 50, 100, and 200 nodes applying tangent sigmoid (tansig) and logistic sigmoid (logsig) activation functions at each layer. The competence of the NN is evaluated with respect to the correlation coefficient R and root-mean-square error (RMSE), which are implemented to measure the correlation and error of the NN outcomes with the VASP results. Based on these evaluations, the optimal combination of parameters is confirmed.[40] For the feature selection using random forest, the max_features, max_depth, and min_samples_leaf are set to 5, 5, and 2, respectively. The critical features that affect the energy are obtained based on the average contribution margin.

3. RESULTS AND DISCUSSION

3.1. Training Results of the Dataset with Equal Acidic Sites of HZSM-5.

3.1.1. ANN Model. The dataset for the ANN model is composed of 2295 sets of HZSM-5 structure−energy data for the analysis of the relationship between HZSM-5 structures and energy, which only includes three acid sites. A total of 2295 × 30,264 data points are contained in the entire data sample, with each structure consisting of 291 atoms producing 30,264 GTSDs. The numbers of structures in the training and test sets are 2197 and 98, accounting for 96 and 4% of the total, respectively. The architecture of the ANN is validated. Table 1 displays the R, RMSE, and training time cost for heterogenous architectures with varying numbers of hidden layers and nodes.

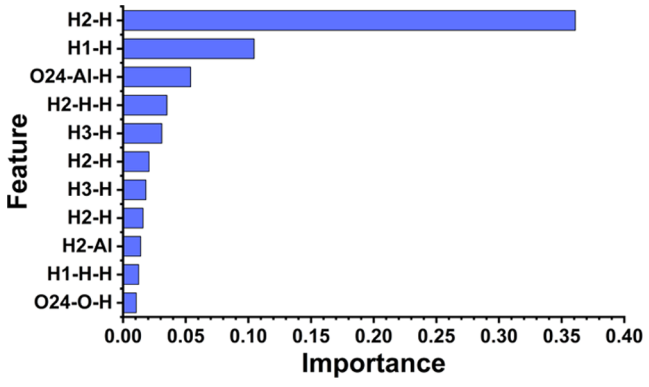


Figure 3. Feature importance ranking.

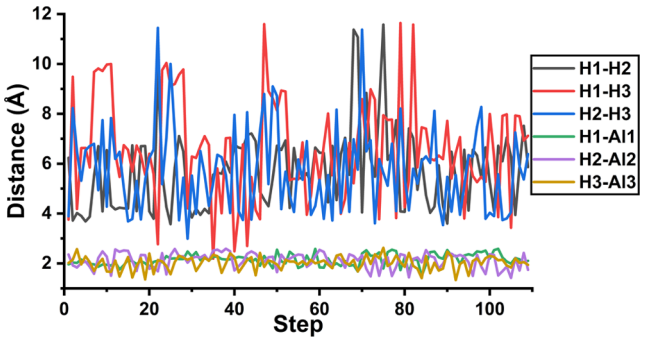


Figure 4. Fluctuations of the distances between H−H and H−Al.

The information provided in Table 1 indicates that the training set R values above 90% are obtained. However, for the test set, poor training accuracy is caused by overfitting the noise in the training set, which can be attributed to an excessively high number of nodes in the test set. The finding is consistent with previous research results.[41,42] Moreover, the training time cost increases with an increase in the number of nodes. Good training effects are acquired for network nodes of 10, 15, and 50 regarding the architectures with relatively high resolution for both the training and test sets; however, only the test set R values of [15, 15] and [10, 10, 10] are greater than 95%. In comparison, the highest R and the smallest RMSE in both the training and test sets are attained by [15, 15].

To assess the influences of activation functions on the training performance, the [15, 15] architecture is kept unchanged, while the activation functions are modified. The training results are summarized in Table 2.

According to Table 2, the R values of training and test sets above 95 and 90% have been achieved by all four activation function combinations as well as the RMSE being in a similar order of magnitude and the training time cost being almost identical. This denotes that the forecasting ability is minimally affected by the type of activation function, and the dataset can be trained by arbitrary combinations. After all aspects are reviewed, [tansig, tansig] is adopted for training the NN model. Consequently, in this study, the [15,15] network architecture with a [tansig, tansig] activation function is chosen. Figure 2 illustrates the training results of the ANN model that implements this configuration.

An ideal effect has been gained through the training of 2197 HZSM-5 structures using the ANN model, with a correlation of R reaching 99.23%, as illustrated in Figure 2a,c. The test set, containing 98 HZSM-5 structures, is employed for testing the network performance utilizing the pretrained ANN model, and the R exceeds 95.0% (Figure 2d). Figure 2e,f demonstrates that the error distribution in the training set is smaller than that in the test set. Figure 2b shows that the energy projected by the ANN has a high degree of coincidence with the energy calculated using VASP, and the RMSE is only 0.020 Ha. A good performance of the energy prediction for HZSM-5 is achieved by the ANN model. In terms of computational time, approximately 200 s is taken for training and forecasting the energy in the ANN, while around 8.8 million seconds are required for performing the task with VASP, indicating an increase in speed of 4 orders of magnitude. Therefore, a feasible approach for fast, exact energy

https://doi.org/10.1021/acsomega.3c06689
ACS Omega 2024, 9, 3392−3400

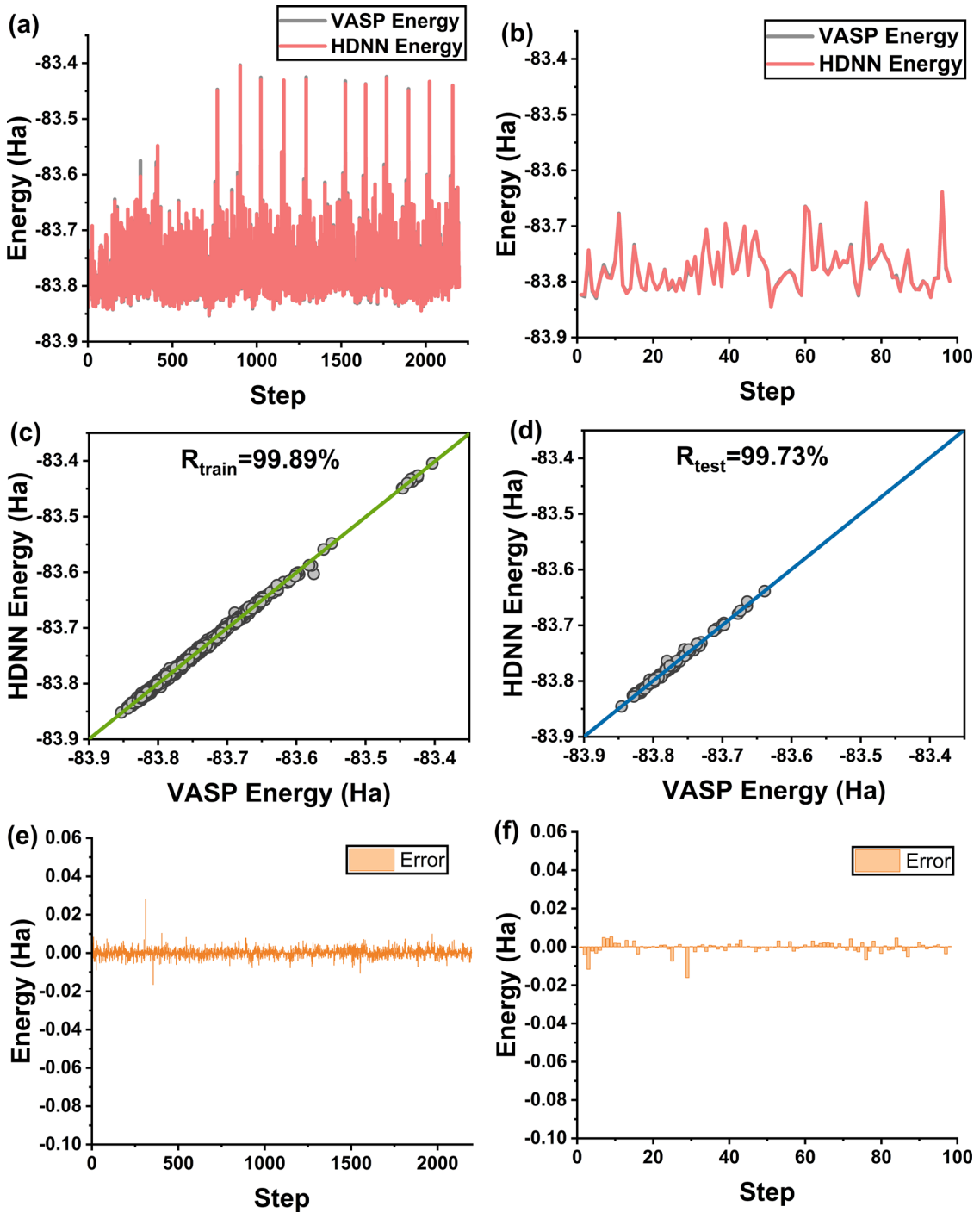


Figure 5. Energy curves of the training set (a) and testing set (b), the correlation between the training set (c) and testing set (d), and the error distribution for the training set (e) and test set (f) of the HDNN model.

Table 3. Training Results of the ANN and HDNN Models

network model	ANN	HDNN
R_{train}	99.23%	99.89%
R_{test}	95.0%	99.73%
RMSE (Ha)	0.020	0.00001
training time cost (s)	100	11147 (3.1 h)

estimation in this data collection operation is provided using the ANN model.

3.1.2. Feature Importance. The relative salience of 30,264 features can be monitored to pinpoint the critical features that play a role in the energy of zeolites. Features with a higher feature contribution are found to contribute more meaningfully to energy. Figure 3 demonstrates the ranking of feature importance, focusing on those features with feature contributions greater than 0.01.

As shown in Figure 3, the main features are the interactions between H atoms with a total feature contribution of 0.6. To further investigate the energy contribution of H and Al to the system, the feature contributions of H−Al and Al−Al are extracted. The results indicate that the contribution of interactions between H−Al and Al−Al is relatively small, at

https://doi.org/10.1021/acsomega.3c06689
ACS Omega 2024, 9, 3392−3400

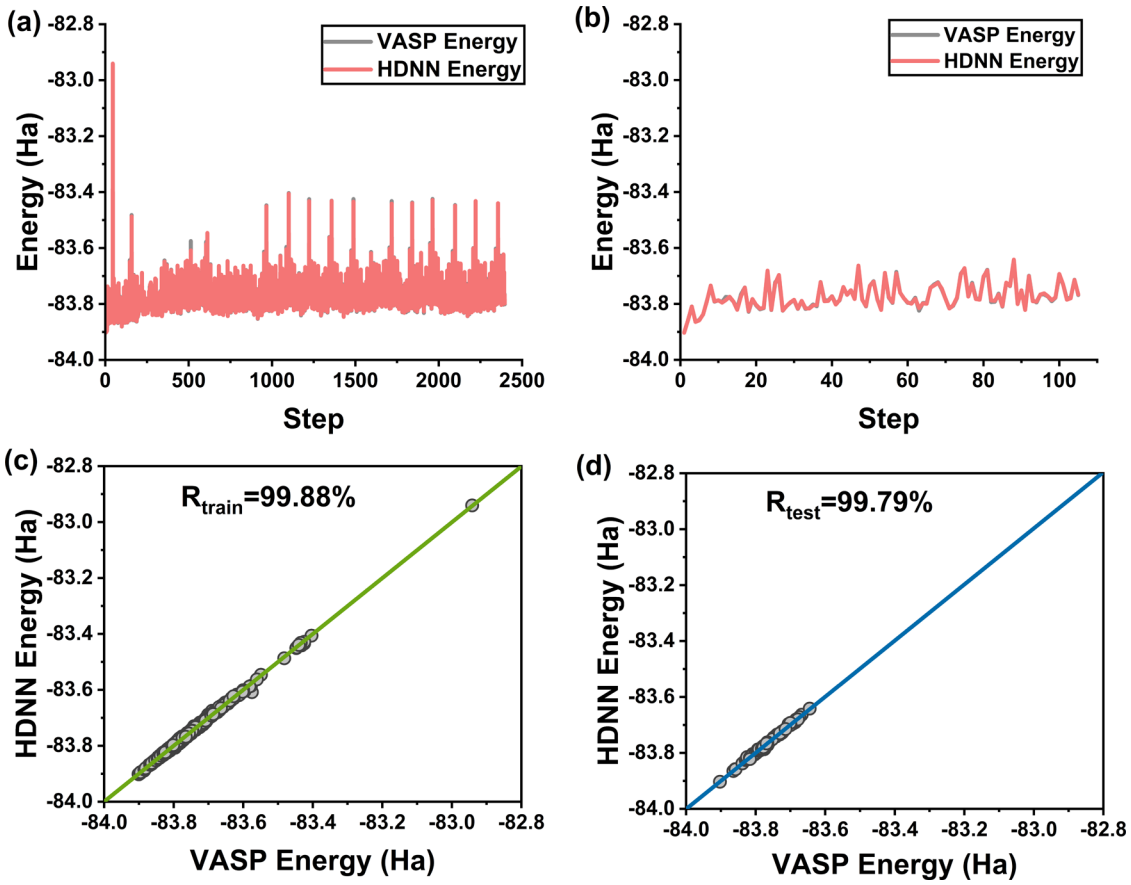


Figure 6. Energy curves of training set (a) and testing set (b) and the correlation between the training set (c) and testing set (d) of the HDNN model with variable numbers of acid sites.

0.0257 and 0.0022, respectively. The locations of acid sites are altered by changing the positions of H and Al in VASP, while maintaining the consistency of the HZSM-5 framework. Significant variation in the positions of H with respect to each other has been detected,[43,44] while those of H and Al remain nearly unchanged. This is likely to be the major reason for the more notable impact of the interaction between H atoms on energy. For quantitative scrutiny, the distances between the H−H and H−Al are extracted from 109 HZSM-5 structures, as illustrated in Figure 4. Minor fluctuations are found in the H−Al distances around 2 Å, while the H−H distances vary within the range of 2−12 Å. It is considered that the variation of H−H distances is the primary factor influencing energy.

3.2. Training Results of the Dataset with Arbitrary Acid Sites of HZSM-5. To compare the differences between the ANN and HDNN, a dataset containing only three acid sites is assigned for training the HDNN with a configuration equivalent to that of the ANN. The forecasts of the HDNN are shown in Figure 5. *R*, RMSE, and the training time cost of the ANN and HDNN are presented in Table 3.

According to Figures 2 and 5, forecasts by the HDNN prove a higher level of coincidence than that of the ANN. As presented in Table 3, *R* values for training and test sets for the HDNN are both higher than those obtained with the ANN: the RMSE training outcomes of the HDNN are much lower than those of the ANN. From Figures 2e,f and 5e,f, it is observed that the error distribution for the HDNN is significantly smaller than that for the ANN for training and test sets. Based on the above-mentioned illustrations, the HDNN is better than the ANN. Although the training time cost of the HDNN is larger, a higher exactness is obtained with the error reduced to 0.00001 Ha. However, the calculation for the HDNN is still significantly more cost-efficient than VASP, improved by nearly 3 orders of magnitude. Integrally considering metrics such as *R*, RMSE, and training time cost, it can be concluded that the predictive capability of the HDNN outperforms that of the ANN. A similar result has also been acquired in the reported research employing other materials, for example, the HDNN is more efficient than the ANN, when considering the system of pressure-induced phase transitions in silicon.[16,21,45] The greater accuracy of the HDNN can be attributed to the utilization of distinct subneural networks, which can predict individual elemental atoms. Learning from different perspectives is facilitated by each network, broadening the scope of molecular features captured by the HDNN and consequently elevating predictive accuracy. In contrast, only a single network is constructed by the ANN, resulting in diminished precision at the molecular level compared with the HDNN.[27,46]

Meanwhile, the ANN is limited by the number of atoms or elements in this research, whereas the HDNN is not.[47] Hence, the energy acquired from 2501 sets of HZSM-5 structures with different numbers of sites is anticipated using the HDNN. The numbers of structures in the training and test sets are 2396 and 105, respectively. The training results are illustrated in Figure 6.

Figure 6a,b demonstrates that the energy predicted by the HDNN is close to the energy calculated using VASP. Additionally, Figure 6c,d indicates that the *R* values of the training set and the test set are 99.88 and 99.79%, respectively.

https://doi.org/10.1021/acsomega.3c06689
ACS Omega 2024, 9, 3392−3400

Meanwhile, the RMSE is only 0.00009 **Ha**, with the training time cost being 12,045 s (3.3 h). With the increase for the original dataset (containing 2295 sets of HZSM-5 structures) with 206 structures, the training time cost is only raised by 0.2 h, which is still 3−4 orders of magnitude more accelerated than that of VASP. In summary, the energy of HZSM-5 structures with varying numbers of sites can be speedily and precisely forecasted by the HDNN. Furthermore, the high accuracy of the HDNN has been documented in a variety of systems, including small molecules, molecular clusters, metal clusters, bulk materials, surfaces, aqueous electrolyte solutions, and solid−liquid interfaces.[48] In summary, the generality of the HDNN is proven.

4. CONCLUSIONS

In this study, the HZSM-5 structures with various distributions of acid sites are utilized to predict energy using machine learning. The optimal parameter combination has been determined via the ANN, and the key features affecting the energy are identified and selected with the aid of a random forest. Additionally, the network performance of both the ANN and the HDNN is assessed on the same data sample. The discussion is conducted on the system-forecasted energy by the HDNN with arbitrary acid sites. The primary conclusions can be drawn as follows:

Based on the *R* and RMSE, the HDNN outperforms the ANN, but the ANN is swifter in terms of training time cost, being nearly 111 times faster than the HDNN. Nevertheless, the HDNN is still approximately 780 times quicker than VASP. Furthermore, good accuracy in energy anticipation of the system with varying numbers of acid sites is exhibited with the use of the HDNN. Taking into account various considerations, superior performance is overall demonstrated by the HDNN.

It has been discovered that the interaction between H−H atoms significantly contributes to the system energy by the random forest feature importance analysis method.

In this work, the advantages and limitations of different NN models have been compared, and in combination with the feature selection approach, key factors affecting energy are screened. New ideas are proposed for the rapid and accurate prediction of the material properties of zeolites and other materials with complex and diverse active site dispersions in nature.

AUTHOR INFORMATION

Corresponding Authors

Bing Zhang − *National Engineering Research Center of New Energy Power Generation, North China Electric Power University, Beijing 102206, People's Republic of China; State Key Laboratory of Alternate Electrical Power System With Renewable Energy Sources and School of New Energy, North China Electric Power University, Beijing 102206, People's Republic of China;* orcid.org/0000-0002-5758-0718; Email: zhangb@ncepu.edu.cn

Bin Hu − *National Engineering Research Center of New Energy Power Generation, North China Electric Power University, Beijing 102206, People's Republic of China; State Key Laboratory of Alternate Electrical Power System With Renewable Energy Sources and School of New Energy, North China Electric Power University, Beijing 102206, People's Republic of China;* orcid.org/0000-0003-2801-9873; Email: binhu@ncepu.edu.cn

Authors

Yang Yang − *School of New Energy, North China Electric Power University, Beijing 102206, People's Republic of China*

Wenming Zhang − *School of New Energy, North China Electric Power University, Beijing 102206, People's Republic of China*

Shengbin Chen − *School of New Energy, North China Electric Power University, Beijing 102206, People's Republic of China*

Xiaogang Wang − *School of New Energy, North China Electric Power University, Beijing 102206, People's Republic of China*

Yuangu Xia − *School of New Energy, North China Electric Power University, Beijing 102206, People's Republic of China;* orcid.org/0009-0005-8257-065X

Ji Liu − *National Engineering Research Center of New Energy Power Generation, North China Electric Power University, Beijing 102206, People's Republic of China; State Key Laboratory of Alternate Electrical Power System With Renewable Energy Sources and School of New Energy, North China Electric Power University, Beijing 102206, People's Republic of China*

Qiang Lu − *National Engineering Research Center of New Energy Power Generation, North China Electric Power University, Beijing 102206, People's Republic of China; State Key Laboratory of Alternate Electrical Power System With Renewable Energy Sources and School of New Energy, North China Electric Power University, Beijing 102206, People's Republic of China;* orcid.org/0000-0002-4340-1803

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c06689

Author Contributions

B.Z. contributed to the conception of the study, wrote the paper together with B.H., Y.Y., W. -M. Z., S. -B. C., X. -G. W., Y. -G. X., and J. L. Y.Y. and W. -M. Z. performed the model training and data processing. Y. Y. performed the data analyses and figure preparation. Q.L. contributed to the constructive discussions and data analyses.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors wish to acknowledge the financial supports from the National Natural Science Foundation of China (62175063, 52276189, 52106241) and the Fundamental Research Funds for the Central Universities (2022YQ002).

REFERENCES

(1) Dědeček, J.; Sobalík, Z.; Wichterlová, B. Siting and distribution of framework aluminium atoms in silicon-rich zeolites and impact on catalysis. *Catal. Rev.* **2012**, *54* (2), 135−223.

(2) Treps, L.; Gomez, A.; de Bruin, T.; Chizallet, C. Environment, stability and acidity of external surface sites of silicalite-1 and ZSM-5 micro and nano slabs, sheets, and crystals. *ACS Catal.* **2020**, *10* (5), 3297−3312.

(3) Li, S.-C.; Lin, Y.-C.; Li, Y.-P. Understanding the catalytic activity of microporous and mesoporous zeolites in cracking by experiments and simulations. *Catalysts* **2021**, *11* (9), 1114.

(4) Soltanian, S.; Lee, C. L.; Lam, S. S. A review on the role of hierarchical zeolites in the production of transportation fuels through catalytic fast pyrolysis of biomass. *Biofuel Res. J.* **2020**, *7* (3), 1217−1234.

(5) Lefebvre, D.; Tezel, F. H. A review of energy storage technologies with a focus on adsorption thermal energy storage processes for heating applications. *Renewable Sustainable Energy Rev.* **2017**, *67*, 116−125.

https://doi.org/10.1021/acsomega.3c06689
ACS Omega 2024, 9, 3392−3400

(6) Lestari, W. W.; Khafidhin, M. A.; Wijiyanti, R.; Widiastuti, N.; Handayani, D. S.; Arrozi, U. S. F.; Kadja, G. T. Novel mixed matrix membranes based on polyethersulfone and MIL-96 (Al) for CO 2 gas separation. *Chem. Pap.* **2021**, *75*, 3337−3351.
(7) Puspitasari, T.; Kadja, G.; Radiman, C.; Darwis, D.; Mukti, R. Two-step preparation of amidoxime-functionalized natural zeolites hybrids for the removal of Pb2+ ions in aqueous environment. *Mater. Chem. Phys.* **2018**, *216*, 197−205.
(8) Puspitasari, T.; Ilmi, M. M.; Nurdini, N.; Mukti, R. R.; Radiman, C. L.; Darwis, D.; Kadja, G. T. M. The physicochemical characteristics of natural zeolites governing the adsorption of Pb2+ from aqueous environment. *Key Eng. Mater.* **2019**, *811*, 92−98.
(9) Mardiana, S.; Azhari, N. J.; Ilmi, T.; Kadja, G. T. Hierarchical zeolite for biomass conversion to biofuel: A review. *Fuel* **2022**, *309*, No. 122119.
(10) Muraoka, K.; Chaikittisilp, W.; Okubo, T. Energy analysis of aluminosilicate zeolites with comprehensive ranges of framework topologies, chemical compositions, and aluminum distributions. *J. Am. Chem. Soc.* **2016**, *138* (19), 6184−6193.
(11) Chen, K.; Horstmeier, S.; Nguyen, V. T.; Wang, B.; Crossley, S. P.; Pham, T.; Gan, Z.; Hung, I.; White, J. L. Structure and catalytic characterization of a second framework Al (IV) site in zeolite catalysts revealed by NMR at 35.2 T. *J. Am. Chem. Soc.* **2020**, *142* (16), 7514−7523.
(12) Janda, A.; Bell, A. T. Effects of Si/Al ratio on the distribution of framework Al and on the rates of alkane monomolecular cracking and dehydrogenation in H-MFI. *J. Am. Chem. Soc.* **2013**, *135* (51), 19193−19207.
(13) He, N.; Ding, Y.-h. Violating the general acidity−activity correlation: Computational evidence in a CpNa-modified HMCM-22 zeolite for ethene protonation. *Microporous Mesoporous Mater.* **2011**, *144* (1−3), 67−73.
(14) Wang, S.; He, Y.; Jiao, W.; Wang, J.; Fan, W. Recent experimental and theoretical studies on Al siting/acid site distribution in zeolite framework. *Curr. Opin. Chem. Eng.* **2019**, *23*, 146−154.
(15) Le, T. T.; Chawla, A.; Rimer, J. D. Impact of acid site speciation and spatial gradients on zeolite catalysis. *J. Catal.* **2020**, *391*, 56−68.
(16) Behler, J. Representing potential energy surfaces by high-dimensional neural network potentials. *J. Phys.: Condens. Matter* **2014**, *26* (18), No. 183001.
(17) Artrith, N.; Morawietz, T.; Behler, J. High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Phys. Rev. B* **2011**, *83* (15), No. 153101.
(18) Tong, Q.; Gao, P.; Liu, H.; Xie, Y.; Lv, J.; Wang, Y.; Zhao, J. Combining machine learning potential and structure prediction for accelerated materials design and discovery. *J. Phys. Chem. Lett.* **2020**, *11* (20), 8710−8720.
(19) Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, *134* (7), No. 074106.
(20) Jose, K. V. J.; Artrith, N.; Behler, J. Construction of high-dimensional neural network potentials using environment-dependent atom pairs. *J. Chem. Phys.* **2012**, *136* (19), No. 194111.
(21) Behler, J.; Martoňák, R.; Donadio, D.; Parrinello, M. Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential. *Phys. Rev. Lett.* **2008**, *100* (18), No. 185501.
(22) Singraber, A.; Behler, Jr.; Dellago, C. Library-based LAMMPS implementation of high-dimensional neural network potentials. *J. Chem. Theory Comput.* **2019**, *15* (3), 1827−1840.
(23) Daeyaert, F.; Ye, F.; Deem, M. W. Machine-learning approach to the design of OSDAs for zeolite beta. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (9), 3413−3418.
(24) Juybar, M.; Khorrami, M. K.; Garmarudi, A. B.; Zandbaaf, S. Determination of acidity in metal incorporated zeolites by infrared spectrometry using artificial neural network as chemometric approach. *Spectrochim. Acta, Part A* **2020**, *228*, No. 117539.
(25) Kwak, S. J.; Kim, H. S.; Park, N.; Park, M.-J.; Lee, W. B. Recent progress on Al distribution over zeolite frameworks: Linking theories and experiments. *Korean J. Chem. Eng.* **2021**, *38* (6), 1117−1128.
(26) Gastegger, M.; Kauffmann, C.; Behler, J.; Marquetand, P. Comparing the accuracy of high-dimensional neural network potentials and the systematic molecular fragmentation method: A benchmark study for all-trans alkanes. *J. Chem. Phys.* **2016**, *144* (19), No. 194110.
(27) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), No. 146401.
(28) Artrith, N.; Behler, J. High-dimensional neural network potentials for metal surfaces: A prototype study for copper. *Phys. Rev. B* **2012**, *85* (4), No. 045439.
(29) Conroy, B.; Nayak, R.; Hidalgo, A. L. R.; Millar, G. J. Evaluation and application of machine learning principles to Zeolite LTA synthesis. *Microporous Mesoporous Mater.* **2022**, *335*, No. 111802.
(30) Li, C.; Vidal-Moya, A.; Miguel, P. J.; Dedecek, J.; Boronat, M.; Corma, A. Selective introduction of acid sites in different confined positions in ZSM-5 and its catalytic implications. *ACS Catal.* **2018**, *8* (8), 7688−7697.
(31) Zhou, M.; Cheng, L.; Liu, B.; Curtiss, L. A.; Assary, R. S. A first-principles investigation of gas-phase ring-opening reaction of furan over HZSM-5 and Ga-substituted ZSM-5. *Ind. Eng. Chem. Res.* **2019**, *58* (33), 15127−15133.
(32) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6* (1), 15−50.
(33) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865.
(34) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **1994**, *50* (24), 17953.
(35) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32* (7), 1456−1465.
(36) MATLAB. https://msca.ncepu.edu.cn/, 2023.
(37) Behler, J. *RuNNer-A neural network code for high-dimensional potential-energy surfaces*. Universität Göttingen 2018.
(38) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(39) LIU, Z. *Machine Learning for Heterogeneous Catalysis: Global Neural Network Potential from Construction to Applications* 2020.
(40) Xu, A.; Chang, H.; Xu, Y.; Li, R.; Li, X.; Zhao, Y. Applying artificial neural networks (ANNs) to solve solid waste-related issues: A critical review. *Waste Manage.* **2021**, *124*, 385−402.
(41) McElroy, P. D.; Bibang, H.; Emadi, H.; Kocoglu, Y.; Hussain, A.; Watson, M. C. Artificial neural network (ANN) approach to predict unconfined compressive strength (UCS) of oil and gas well cement reinforced with nanoparticles. *J. Nat. Gas Sci. Eng.* **2021**, *88*, No. 103816.
(42) Jin, L.; Kuang, X.; Huang, H.; Qin, Z.; Wang, Y. Study on the overfitting of the artificial neural network forecasting model. *ACTA METEOROLOGICA SINICA-ENGLISH EDITION-* **2005**, *19* (2), 216.
(43) Zhai, D.; Liu, Y.; Zheng, H.; Zhao, L.; Gao, J.; Xu, C.; Shen, B. A first-principles evaluation of the stability, accessibility, and strength of Brønsted acid sites in zeolites. *J. Catal.* **2017**, *352*, 627−637.
(44) Li, S.; Li, H.; Gounder, R.; Debellis, A.; Müller, I. B.; Prasad, S.; Moini, A.; Schneider, W. F. First-principles comparison of proton and divalent copper cation exchange energy landscapes in SSZ-13 zeolite. *J. Phys. Chem. C* **2018**, *122* (41), 23564−23573.
(45) Behler, J.; Martoňák, R.; Donadio, D.; Parrinello, M. Pressure-induced phase transitions in silicon studied by neural network-based metadynamics simulations. *Phys. Status Solidi B* **2008**, *245* (12), 2618−2629.
(46) Chen, L.; Zhang, X.; Chen, A.; Yao, S.; Hu, X.; Zhou, Z. Targeted design of advanced electrocatalysts by machine learning. *Chin. J. Catal.* **2022**, *43* (1), 11−32.

https://doi.org/10.1021/acsomega.3c06689
ACS Omega 2024, 9, 3392−3400

(47) Gastegger, M.; Marquetand, P. High-dimensional neural network potentials for organic reactions and an improved training algorithm. *J. Chem. Theory Comput.* **2015**, *11* (5), 2187−2198.
(48) Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem., Int. Ed.* **2017**, *56* (42), 12828−12840.

https://doi.org/10.1021/acsomega.3c06689
ACS Omega 2024, 9, 3392−3400