

DeepNCI: DFT Noncovalent Interaction Correction with Transferable Multimodal Three-Dimensional Convolutional Neural Networks

Wenze Li, Donghan Wang, Zirui Yang, Huijie Zhang, LiHong Hu,* and GuanHua Chen*



Cite This: <https://doi.org/10.1021/acs.jcim.1c01305>



Read Online

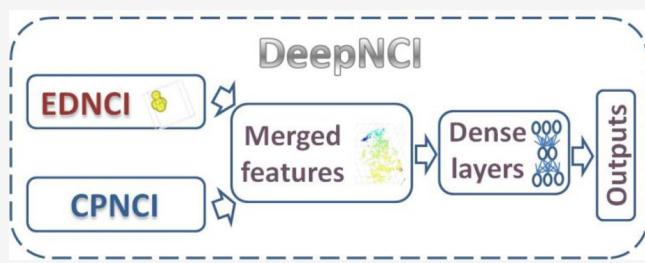
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: A multimodal deep learning model, DeepNCI, is proposed for improving noncovalent interactions (NCIs) calculated via density functional theory (DFT). DeepNCI is composed of a three-dimensional convolutional neural network (3D CNN) for abstracting critical and comprehensive features from 3D electron density, and a neural network for modeling one-dimensional quantum chemical properties. By merging features from two networks, DeepNCI is able to reduce the root-mean-square error of DFT-calculated NCI from 1.19 kcal/mol to ~0.2 kcal/mol for a NCI molecular database (>1000 molecules). The representativeness of the joint features can be visualized by t-distributed stochastic neighbor embedding (t-SNE), where they can distinguish categorized NCI systems quite well. Therefore, the fused model performs better than its component networks. In addition, the 3D CNN takes electron density as inputs that are in the same range, despite the size of molecular systems, so it can promote model applicability and transferability. To clarify the applicability of DeepNCI, an application domain (AD) has been defined with merged features using the *K*-nearest-neighbor method. The calculations for external test sets are shown that AD can properly monitor the reliability for a prediction. The model transferability is tested with a small database of homolysis bond dissociation energy including only dozens of samples. With NCI database pretrained parameters, the same or better performance than the reported results is achieved by transfer learning. This suggests that the DeepNCI model is transferable and it may transfer to other relative tasks, which possibly can resolve some small sampling problems. The source code of DeepNCI can be freely accessed at <https://github.com/wenzelee/DeepNCI>.



INTRODUCTION

Intermolecular interactions such as protein folding, drug binding, and nanomaterial self-assembling are frequently dominated by noncovalent interactions (NCIs).^{1–5} They span a wide range of weak interactions, including hydrogen bonding, dipole–dipole interactions, steric repulsion, and London dispersion, etc. Experimental measurements of NCIs are highly demanding,^{6–8} especially for large molecules containing abundant NCIs; therefore, effective computational tools are often resorted to, in particular, high accurate first-principles methods. Apparently, NCI calculations have more complexity than covalent bonds, because of its complexity and diversity. Moreover, intrinsic mechanisms of some NCIs are still unclarified and explicit formulation descriptions for them are not yet available. When confronting NCI problems, some considerations, such as long-range interaction, correlation, dispersion, and polarization, should be involved in the functionals of computational methods.⁹ The nonvariational method with correlation of the electronic motions, CCSD(T), has been termed the gold standard of NCI computation.¹⁰ However, the main obstacle that obstructs widespread applications of CCSD(T) in practice, has been the steeply increasing computational effort $> O(N^6)$ with the number of molecular orbitals N .^{10–12} In past decades, the density

functional theory (DFT) method has become one of the most popular first-principles methods in molecular studies, because of their efficiency and economy. However, they are deficient on NCI computations. Conventionally, in DFT methods, parametrization, dispersion correction, or long-range correction are often adopted in functionals to perform NCI calculations.^{13,14} Whereas improving functionals is confined with the capability of computational methods, massive improvements cannot be guaranteed and computation is still costly. A recently emerged alternative, machine learning correction, can circumvent intractable problems in first-principles methods, such as effective approximations, limited basis set, and explicit function forms.^{15–18} Therein, machine learning models can dramatically improve calculation accuracies to a high level of first-principles (MP2, CCSD, etc.) with input features by low-level calculations, e.g., DFT or Hartree–

Special Issue: Computational Chemistry in Asia

Received: October 28, 2021

Fock with small basis sets such as STO-3G and 6-31G*.^{19–23} Especially in our proposed machine learning correction model,¹⁹ it takes a calculated target value (e.g., DFT-calculated NCI in this study) by first-principles methods as a primary descriptor that contains most of the physical essence of the target, which makes the accuracy improvement for the low-level calculation easy to attain. The molecular descriptors obtained by first-principles methods have physical meanings and are apt to construct high-capacity machine-learning models with better interpretability. Therefore, cooperating first-principles calculations with machine-learning algorithms probably can achieve high accuracy and efficiency simultaneously.^{24–29} Recently, with the rapid development of artificial intelligence and enormous applications of first-principles methods, machine learning has started to play a more and more important role in computational chemistry. However, the machine learning model performance is heavily dependent on the quality of input features that are mostly obtained by feature engineering.³⁰ Nevertheless, feature design and selection are ambiguous tasks that count on special expertise and means to find the complete set of significant features.^{20,21,30} Fortunately, the deep learning algorithm is determined to be an effective solution.

In 2012, it was a milestone that Hinton's group won the Merck Kaggle challenge (<https://www.kaggle.com/c/MerckActivity>) using their deep learning models, which opened a new chapter of applications using deep learning on predicting activity and property for chemical compounds. Deep learning is a representative learning, and its applications have been soaring in various areas in the past decade.^{31–39} It arises with its specialty of transforming the original raw input features to distributed representation through multilayer networks. Because of this signal extraction capability, it can excavate representative features and reveal their internal association automatically, without human intervention. Thus, it can make objective decisions based on the abstracted feature and also avoid the troublesome feature engineering in traditional machine learning. Nowadays, deep learning has already exhibited prominence on quantum chemical calculations. Yao and Parkhill³⁷ used the 3D electron density of small molecules, rather than 2D molecular fingerprints or physical chemical properties, as the input feature, and developed a 3D convolutional neural network model to predict the Kohn–Sham kinetic energy of hydrocarbons. Rathi et al. presented a deep graph convolutional neural network model, trained on electrostatic potential surfaces derived from high-quality first-principles calculations, that rapidly generates ESP surfaces for ligands.⁴⁰ Torg et al. proposed a graph-convolutional neural network (Graph-CNN) framework for predicting positive or negative interactions between targets and ligands.⁴¹ Zhou and co-workers employed a three-dimensional convolution neural network (3D CNN) as the universal exchange-correlation potential, and the quasi-local electron density is used as the descriptors.⁴² It was also employed to determine the NCI of a helium dimer.⁴² An experiment showed the Graph-CNN models with fine prediction accuracy, outperformed other models, and achieved comparable performance to 3D CNN protein–ligand scoring.⁴¹ Lim recently proposed a novel deep learning approach, using a graph neural network for predicting noncovalent interactions between drugs and protein targets.⁴³ They extracted the graph feature of intermolecular interactions directly from the 3D structure on the protein–ligand binding

pose to learn key features for accurate predictions of drug–target interactions. The method is superior to other deep learning models on the same task.⁴³ Although many deep neural networks contributed to NCI, they mainly have been focusing on classification decisions.^{41,43} The regression work with deep learning for predicting accurate NCI so far has been discussed little.

In this study, we explored high accurate NCI regression deep learning model with multidimensional descriptors. Three general modeling problems are mainly addressed: universal molecular representation, high prediction accuracy, and transferability. (1) To improve the model generalization, electron density (size-independent) is taken as raw input features of 3D CNN, which can extract the representative features from electron density cube. (2) To obtain high accuracy, a multimodal deep learning architecture named DeepNCI was established for NCI correction, which fused two different types of modalities, 3D CNN and a neural network (NN) model, dealing with electron density cube and chemical property descriptors, respectively. With multimodality, DeepNCI can perform both feature extraction for 3D electron density cube and descriptor transformation for 1D chemical properties, thus gave high accurate NCI prediction. (3) Transferability means extended applications for a model. So the transferable property has been tested for DeepNCI on a homolysis bond dissociation energy database. As far as we know, the multimodal with deep architecture for the NCI regression has not been reported.

MATERIALS AND METHODS

Datasets. When turning to benchmark databases of NCI, the eminent works by Professor Hobza's group should be highly appreciated. In the modeling, the NCI benchmark databases $S_{66} \times 8$,⁴⁴ $S_{22} \times 5$,⁴⁵ $X_{40} \times 10$ ⁴⁶ are used. The 1038 molecules in the databases are classified into four categories, according to the dominant NCIs that are presented as dispersion, hydrogen bonding, mixed complexes, and halogen interactions.

The $S_{66} \times 8$ dataset consists of 66 molecular systems, herein each of them is treated at eight different intermolecular distances, which allows an evaluation for the long-range behavior of the investigated methods. Geometries were constructed by scaling the equilibrium intermolecular distance in the complexes by a factor of 0.9, 0.95, 1.0, 1.05, 1.1, 1.25, 1.5, and 2.0, starting from the MP2/cc-pVTZ geometry. The $S_{22} \times 5$ dataset consists of 22 molecular systems, and the geometries of noncovalent complexes from the S_{22} dataset were displaced along the intermolecular axis, forming one shortened and three elongated (0.9, 1.2, 1.5, and 2.0 times of the equilibrium intermolecular distance) structures for each molecule. The $X_{40} \times 10$ dataset features 10 points (equilibrium distance r_e , four compressed dimers at $0.80r_e$, $0.85r_e$, $0.90r_e$, and $0.95r_e$, and five stretched dimers at $1.05r_e$, $1.10r_e$, $1.25r_e$, $1.50r_e$, and $2.00r_e$) along the dissociation curve of each molecule in the 40 complexes of the X_{40} dataset, which covers electrostatic interactions, London dispersion, hydrogen bonds, halogen bonding, halogen–π-interactions, and stacking of halogenated aromatic molecules. In our previous works, we tested various DFT methods, and in terms of accuracy, efficiency, and stability, M062X is considered as the best choice for computing quantum chemical descriptors (DFT NCI value, electron density and other quantum chemical properties) among tested DFT methods.^{20,21} Therefore, single-

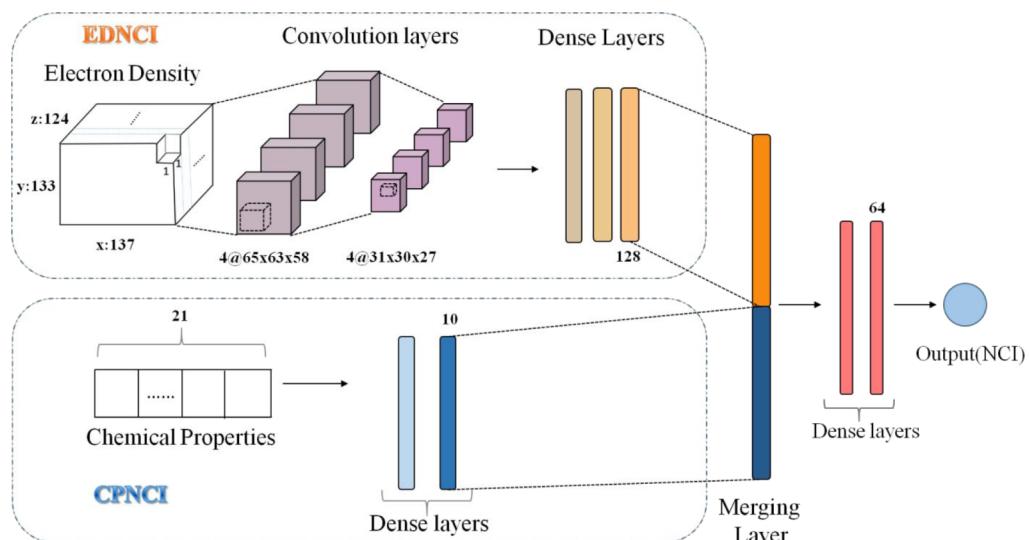


Figure 1. DeepNCI framework consisting of two subnetworks EDNCI and CPNCI.

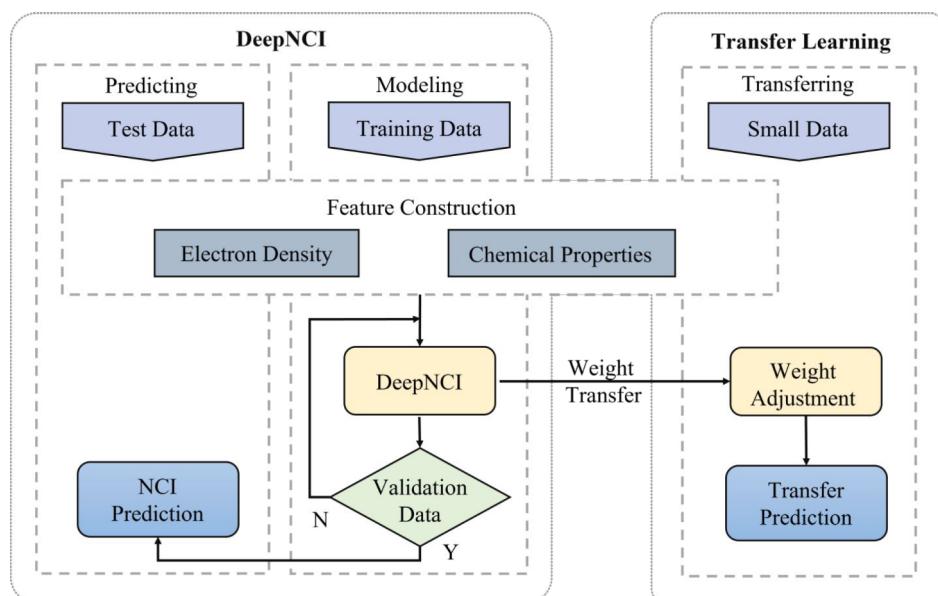


Figure 2. Flowchart of DeepNCI model and transfer learning scheme: (left) the training procedure of DeepNCI and (right) the transfer learning strategy.

point calculations of these 1038 molecules are performed by the Gaussian09 program package,⁴⁷ using DFT M062X method with the 6-31G* basis set under vacuum.

For building an unbiased and generalized model, a hybrid distance sample set partition based on joint x-y distances (HSPXY)⁴⁸ is leveraged to divide the entire database into a training set (800 molecules), an independent test set (214 molecules) and a validation set (24 molecules). The validation set is used to validate the model and adjust the parameters. The independent test set evaluates the prediction and generalization ability of models.

Molecular Features. Two types of features of NCI molecules calculated by the DFT (M062X/6-31G*) method are exploited as model inputs that are 3D electron density and quantum chemical properties.

Electron Density. The convolutional neural networks have been trained to reproduce the Kohn–Sham kinetic energy or the exchange-correlation potential from the electron den-

sity.^{37,42} Similarly, NCIs may induce the changes of electron density, leading to a more-diffused distribution. Since molecular structures are three-dimensional, the corresponding electron density is a cubic matrix. The volume of density cubes ($\sim 80^3$ voxels) varies according to the size of molecules. To construct the deep learning model, the electron density cubes of all molecules for the inputs must be normalized with the largest cubic box [$x:137$, $y:133$, $z:124$] by taking the maximum in the x,y,z directions of all the samples and zero padding is employed in the vacant space.

Molecular Properties. In refs 20 and 21, 43 molecular descriptors consisting of 25 quantum chemical and 18 structural properties were adopted as the inputs in the NCI models. Herein the 18 structural descriptors were not employed in this study, because the structural information on S66 \times 8, S22 \times 5 and X40 \times 10 is the same as their respective source equilibrium structures in S66, S22, and X40 benchmark databases. In addition, four quantum chemical

properties (the total charge on monomer A or B, the polarizability of each monomer) were discarded for their locality. The list of chemical descriptors used in this study is given in Table S1 in the Supporting Information(SI).

Deep Architecture Construction. The proposed deep architecture DeepNCI is represented in Figure 1. It is a fused neural network that consists of two subnetworks that deal with electron density in convolution layers and quantum chemical descriptors by neural network dense layers, respectively; their outputs are merged into a fully connected network for NCI regression. All the hyper-parameters in the network are listed in Table S2 in the Supporting Information.

In one subnetwork named as EDNCI (NCI CNN model with electron density inputs), there are two convolutional layers and three max-pooling layers to process the 3D electron density cube. EDNCI can extract important information from sparse electron density in cubes. The obtained representative features are taken as inputs of fully connected dense layers. The inputs of 3D electron density cube ($137 \times 133 \times 124$) are highly dimensional, therefore two convolutional layers with pooling are used to extract features and reduce the dimensionality, which lowers the risks of the overfitting problem. For this purpose, Drop-out strategy is also adopted.

In the other subnetwork called CPNCl(NCI NN model with chemical property inputs), only a dense layer is used to generate their representations, because we found the fully connected neural network rather than CNN is more suitable for nonsequential descriptors. The output features (EDNCI 128 and CPNCl 10) generated by two subnetworks are merged into inputs of fully connected dense layers.

The experiments are run on the Tensorflow 3.5, with a NVIDIA GeForce GTX 1080 Ti GPU. Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$ is used to optimize the parameters at each iteration. The learning rate varies over the course of training, according to the following equation:

$$\begin{aligned} &\text{decayed learning rate} \\ &= \text{learning rate} \times \text{decay rate}^{\wedge}(\text{global_step}/\text{decay_steps}) \end{aligned}$$

where the learning rate = 10^{-4} , global_step = 300, decay_steps = 70, and decay_rate = 0.97.

Generally, several hours are needed to construct an optimal training model. Because of GPU acceleration, predicting NCIs for a molecule only costs tens of seconds.

The flowchart of modeling is shown in Figure 2. In feature construction, two types of features are used for all the training, validation, and test samples. During training, the training set was used to train the parameters in EDNCI, CPNCl, and DeepNCI, and the validation set was used to evaluate the model and adjust the parameters. On the prediction, the performance of model was evaluated by the test set. The evaluation parameters, root-mean-square error (RMSE), mean absolute error (MAE), and predictive squared correlation coefficient (Q^2) are defined by eqs 1–3 in the Supporting Information. The transfer learning procedure is the same as DeepNCI, except for data with a new task for related/similar targets. Moreover, it uses the weights of the well-trained DeepNCI as the initialization for training with the same framework and hyperparameters.

Transferability. Transfer learning refers to the learning model that has been trained in previous tasks, which can be migrated to target data directly or in a continuous learning manner, through parameters, features, and samples. It has been

widely applied on massive image classification issues, and excellent performance on small-scale data has been obtained. Generally, its advantages are (1) using labels in other similar tasks that are easy-to-obtain tags, when the labels of the target tasks are scarce; (2) offering a great starting point for modeling a new database, which can dramatically accelerate the learning process.

In our previous study, a homolysis bond dissociation energy (HBDE) database containing 92 samples was reported. (The molecule structures are given in Table S3 in the Supporting Information.)⁴⁹ It is a very small database, which is impossible to be adopted for deep learning methods, since deep learning networks are required to fit enormous parameters. To learn the transferability and applicability of DeepNCI on small-scale databases, we utilized the pretrained network based on NCI data for the small HBDE database that contains a series of nitric oxide (NO) carrier molecules.⁴⁹ Herein, the parameter transfer mode is applied, and the detailed framework of transfer learning is shown in Figure 2. First, the entire DeepNCI network with pretrained parameters is transferred to the HBDE data. Then, the model is fine-tuned by the HBDE training data and makes a prediction for the test set. In Figure 2, the left side shows the pretraining model, DeepNCI, obtained by training NCI data, and the right side shows the transferring model for the HBDE data. The same color components in Figure 2 represent equivalent network modules of DeepNCI model transferring to the target HBDE model. In addition, the different color components stand for the specific input and output layers of the transfer learning model.

RESULTS AND DISCUSSIONS

DeepNCI Performance. DeepNCI is a regression model including two types of networks for the prediction of NCI

Table 1. Evaluation Parameters of DFT, Various Machine Learning Models on the Test Set

model	root mean square error, RMSE	mean absolute error, MAE	predictive squared correlation coefficient, Q^2
DFT	1.13	0.82	0.887
EDNCI	0.55	0.35	0.973
EDNCI +NCl _{DFT}	0.26	0.18	0.994
CPNCl	0.38	0.28	0.987
DeepNCI	0.19	0.14	0.997

values. The training curve is shown in Figure S1 in the Supporting Information, where the loss becomes steady after ~ 150 epochs. Comparisons of DeepNCI performance with its two subnetworks (EDNCI and CPNCl) and DFT-only calculations were performed, based on the test set, and their evaluation parameters are given in Table 1. Benefiting from the multimodal synergies, DeepNCI model obviously achieved better performance than both unimodal subnetworks, CPNCl and EDNCI. The RMSE and MAE of DeepNCI were reduced by 83% and 84% from DFT calculations, respectively, which are just around half of any subnetworks errors. It suggests that the merged strategy effectively unifies the advantages of the subnetworks. However, EDNCI performed worse than CPNCl, which might result from the primary descriptor, i.e., the calculated NCI by M062X/6-31G* used in the CPNCl model. As we proposed in the correction model,^{19–22} the DFT-calculated target is taken as the primary descriptor, which

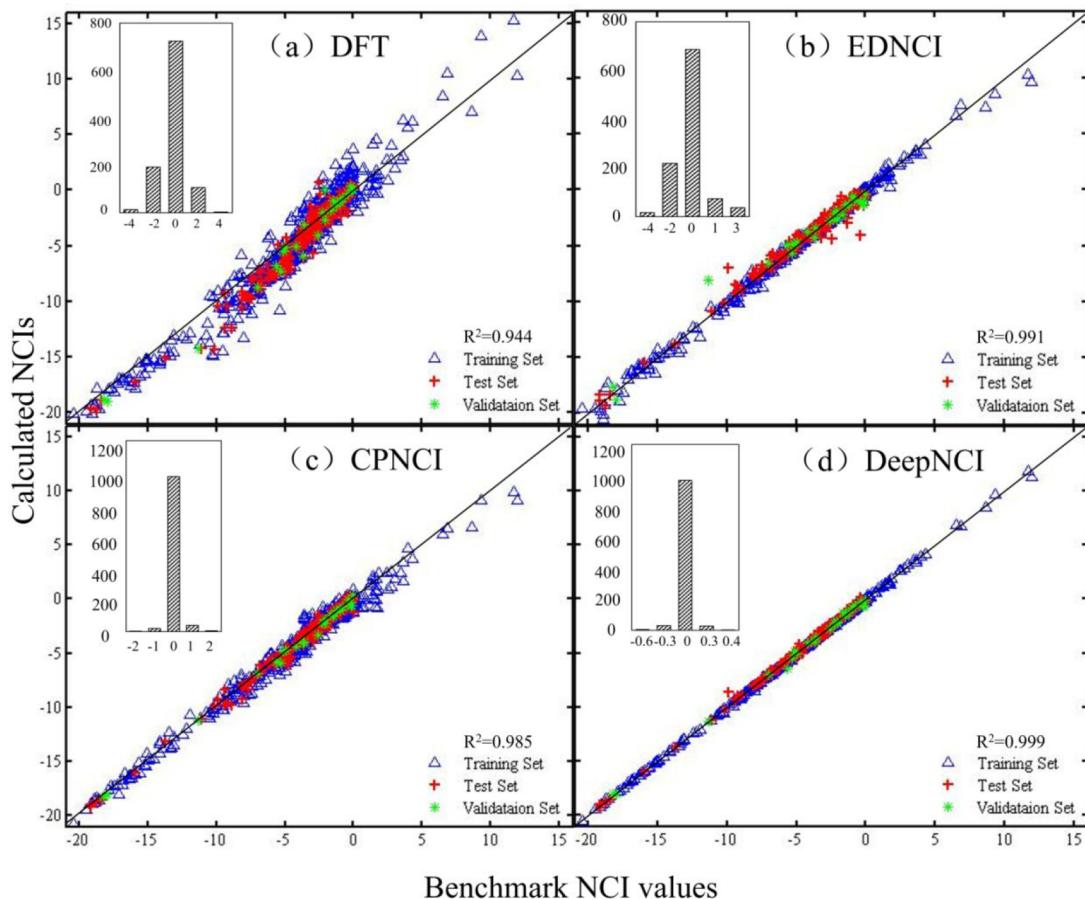


Figure 3. Plots of the calculated versus benchmark interaction energies for the training set, test set, and validation set in different models.

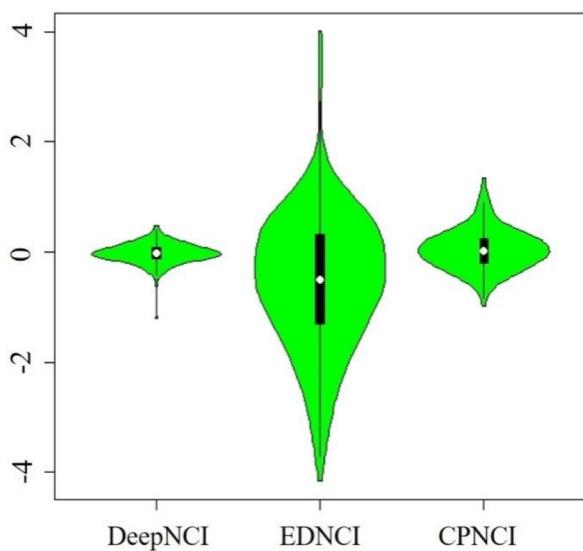


Figure 4. Violin plots of prediction errors of NCIs in different models. The errors of the test set are calculated with $\text{NCI}_{\text{benchmark}} - \text{NCI}_{\text{Model}}$.

makes improvement/correction easy, since all the essential information on the NCI target has been included. Therefore, we add only this key feature into the EDNCI model (EDNCI + NCI_{DFT} in Table 1) to determine how the model performs. Unsurprisingly, the RMSE of EDNCI+ NCI_{DFT} has been remarkably improved to 0.26 kcal/mol after adding the key feature. Another test, CPNCI modeling without the key

Table 2. Results on reported 121 molecules by different models

method	root mean square error, RMSE	mean absolute error, MAE	predictive squared correlation coefficient, Q^2
GRNN ²¹	0.50	0.35	0.99
Hete-SE ²⁰	0.31	0.20	0.99
DeepNCI	0.12	0.08	0.99

feature, is also performed, where the RMSE goes up to ~ 1.1 kcal/mol. This also illustrates that abstracted features from the electron density cube are more representative than quantum chemical property descriptors without the primary descriptor, and either subnetwork plays an important role in the fused multimodal.

To learn the significance of the primary descriptor, the permutation error method is performed for all quantum chemical properties, and the histogram of RMSE increase induced by randomly changing the order of descriptors is shown in Figure S2 in the Supporting Information. In the figure, the RMSE increase of DFT calculated NCI is much larger than other descriptors, which indicates its prominent significance. The second important group of quantum chemical descriptors includes polarizability, potential energy, kinetic energy, $E_{\text{LUMO}+1}$, etc. Note that the polarizability is one of driving forces for changing electron density shape/distribution; and most other important descriptors are related to the electron energy. These may imply certain electronic characteristics are some major factors to predict NCIs. The scatter plots

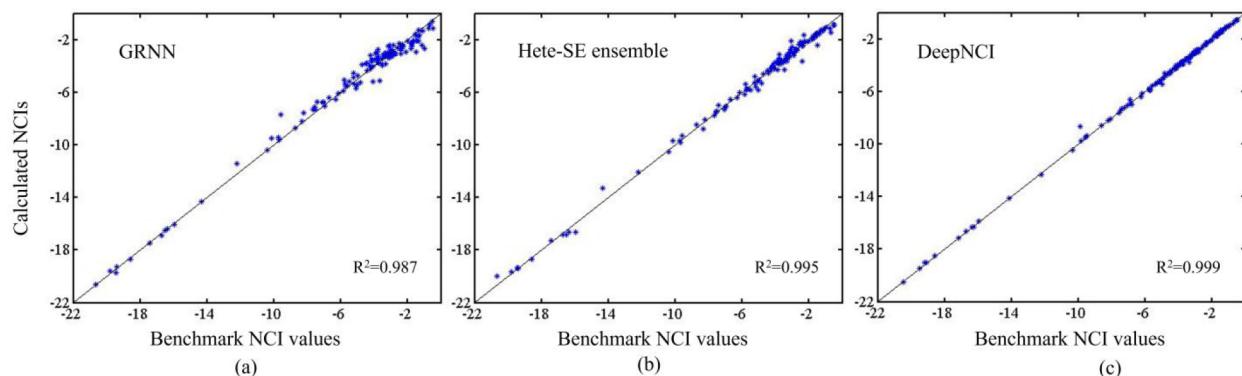


Figure 5. Plots of calculated values versus benchmark interaction energies for reported 121 molecules by a series of models: (a) GRNN, (b) Hete-SE ensemble, and (c) DeepNCI.

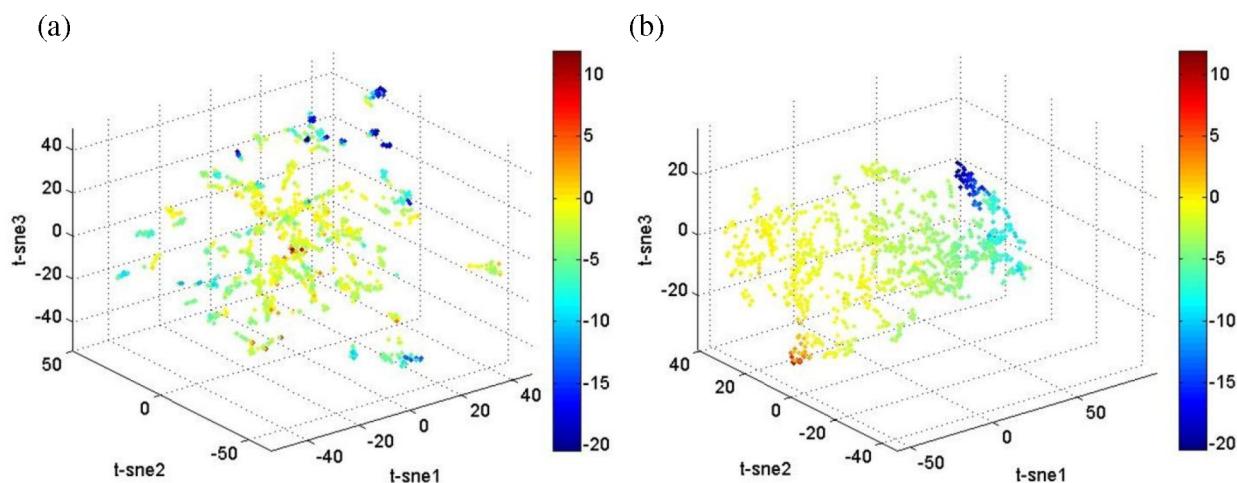


Figure 6. T-SNE visualization of two states of obtained features in three dimensions: (a) 3D CNN extracted feature layer and (b) merged feature layer.

for the calculated versus the benchmark NCIs values by various models are shown in Figure 3. It is obvious that the calculated values by multimodal DeepNCI agree with the benchmark values better than those by both subnetwork models, which exhibited DeepNCI superiority. The NCI prediction values for the test set are listed in Table S4 in the Supporting Information.

The violin plots of predictive errors for DeepNCI, EDNCl and CPNCl are shown in Figure 4. They begin with a box plot; then, a rotated kernel density plot is added to each side of the box. It can be seen that the error of DeepNCI only spans a narrow range of errors (-1.2–0.5), almost all of the test samples achieve the chemical accuracy. This shows the multimodal DeepNCI has been strengthened by fusing two machine learning models. In addition, an uncertainty 0.23 (0.07) kcal/mol has been estimated by run 30 trainings with a randomized validation set.

NCI Model Comparisons. We compared DeepNCI with GRNN (single machine learning model)²⁰ and Hete-SE (ensemble learning model).²¹ The comparison results for the reported 121 molecular database are shown in Table 2. The results of GRNN and Hete-SE were taken directly from refs 20 and 21. Moreover, the training and test datasets in DeepNCI were set the same as that in our reported experiments. From the evaluation parameters in Table 2, it is noted that the NCI improvement for 121 molecules by DeepNCI is superior to

single and ensemble models (RMSE: 0.12 vs 0.31 or 0.50 kcal/mol).

The scatter plots of the calculated versus the benchmark NCIs values for the reported 121 molecules by three models are shown in Figure 5. It has been shown that the results calculated by DeepNCI are obviously better, compared to that observed by GRNN and Hete-SE. In Figure 5c, the calculated NCI values by DeepNCI are closer to the $y = x$ straight line than that by either Hete-SE or GRNN in Figures 5a and 5b. The correlation coefficient (R^2) between predicted and benchmark NCIs on the lower right corner increases with the accuracies of models. Again, this shows that deep learning is more capable than simple and ensemble machine learning models, which indicates representative molecular features might have been extracted by DeepNCI. Recently, data sets S66 \times 8 and X40 \times 10 used to model NCI were reported in ref 27, where the smallest MAEs of S66 \times 8 and X40 \times 10 datasets by their NCI corrected schemes are 0.1 and 0.18 kcal/mol, respectively. In our model, MAEs for these two datasets are 0.06 and 0.08 kcal/mol, respectively. This suggests DeepNCI may be a fine model for NCI corrections.

Feature Visualization. To explore extracted features by DeepNCI, the t-SNE was used to visualize the high-dimensional features through a three-dimensional map. The visualization of features is shown in Figure 6, and the color bar corresponds to NCI values. It can be seen that all samples are mixed together in Figure 6a, when plotting with CNN

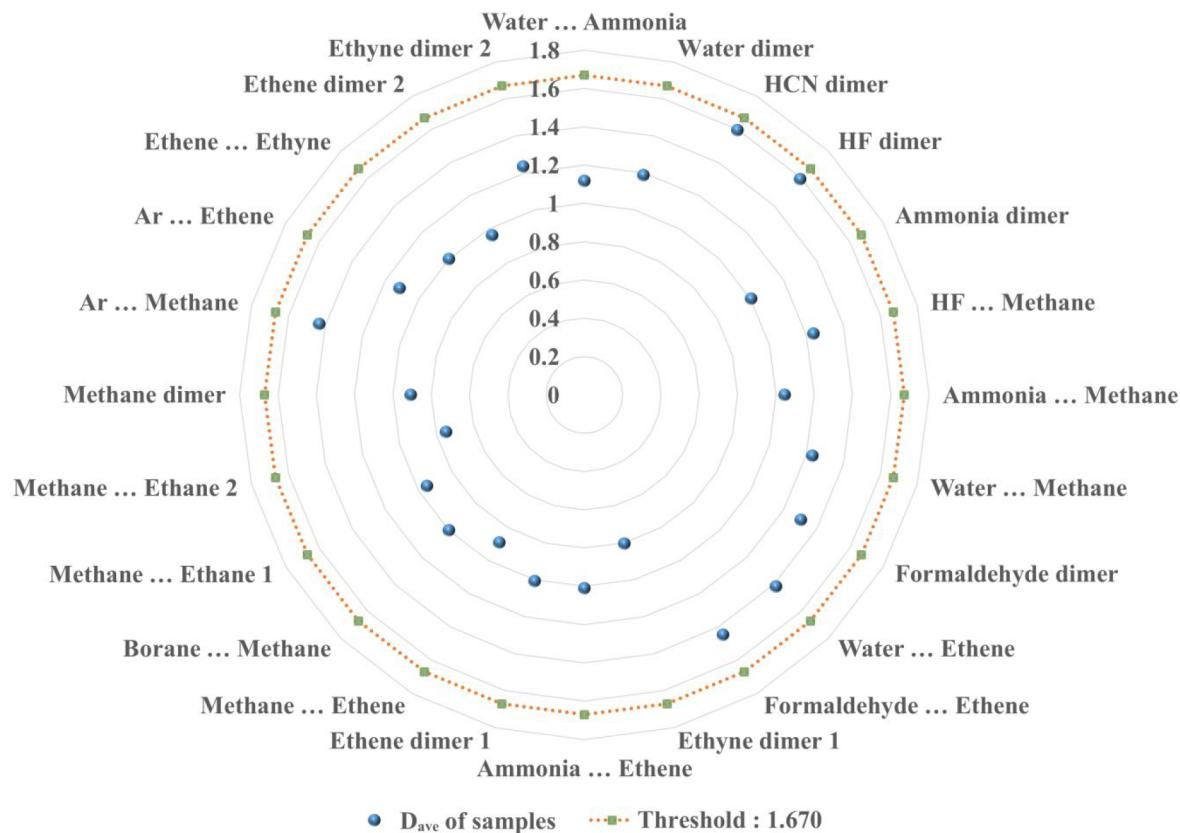


Figure 7. Radar plot of DeepNCI application domain illustrated by the external test set (A24). The orange dotted line indicates the threshold (1.67) and blue dots are the D_{ave} values of test samples.

Table 3. Results on HBDE Data Based on Various Models

method	RMSE of HBDE data (kcal/mol)
DFT	5.31
DeepNCI	0.24
MIV-BPNN	0.33
GP-GRNN	0.31

extracted features from the electron density cube. Whereas, in Figure 6b, after DeepNCI merged multilayer processing, formed features can explicitly separate NCI samples to several clusters. It implies that the deep architecture DeepNCI may extract important representations from input features and

merging them can further enhance the representative, thus model predictive capability. That is to say, DeepNCI is able to generate significant features to recognize molecules with similar NCI or having close NCI values, and distinguish molecules dominating with different types of NCIs.

Application Domain. The above analyses show that DeepNCI is capable of giving highly accurate NCI predictions. However, with limited data, it has application constraints. Therefore, to reveal DeepNCI applicability, an application domain (AD) has been defined by the K -nearest-neighbor (KNN) method. Via KNN, the Euclidean distances between all training samples and their K -nearest neighbors are calculated by merged features (138) in the merged layer; herein, K is

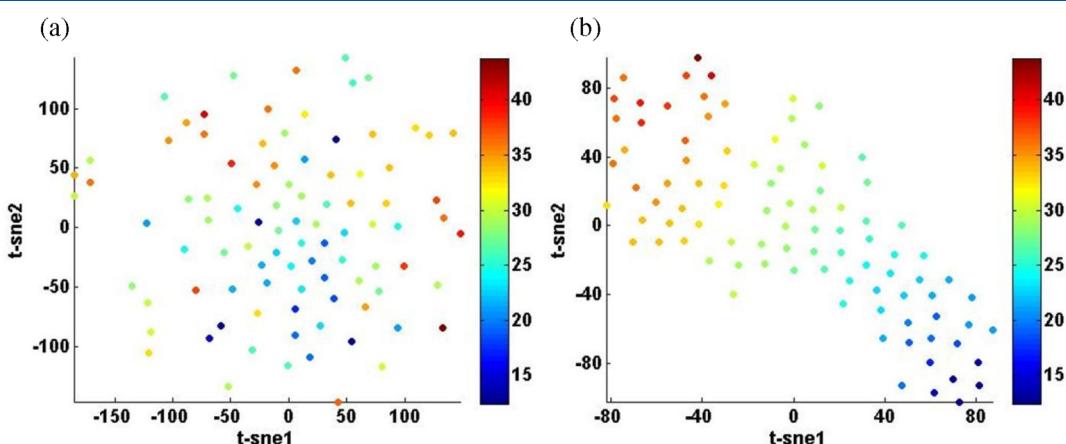


Figure 8. T-SNE visualization of (a) CNN extracted features and (b) merged features on the HBDE dataset.

optimized to 5. The average of five distances for each sample is expressed as D_{ave} . The 95th percentile of all D_{ave} is taken as the domain threshold. If the D_{ave} value of a test sample is equal or less than the threshold, it is considered to be inside the AD. The prediction uncertainty increases as data go beyond the threshold. The threshold (1.67) of DeepNCI is illustrated in Figure 7. With a defined AD, the D_{ave} value of two external test sets (A24⁵⁰ and neutralized molecules in SCAI; Representative amino acid side chain interactions in proteins)⁵¹ is calculated. For clarity, only the D_{ave} of A24 is shown in Figure 7. It is noted that all the D_{ave} values of A24 samples are inside the AD, and the prediction errors (shown in Table S5 in the Supporting Information) for external datasets are in similar accuracy as that of the test set. Calculated NCIs and D_{ave} of SCAI are shown in Table S6 and Figure S3, respectively, in the Supporting Information, where it shows some molecules outside of AD have large prediction errors. This suggests that the KNN-defined AD can reasonably guide the user to employ the DeepNCI model. In addition, we must say that the predictable range of NCI values is approximately -20–0 kcal/mol for neutralized molecules, corresponding to the training data. Moreover, the current version of programs only can compute samples with the cubic box volume within $x:137 \times y:133 \times z:124$. We have tested inputs with an enlarged cubic box to $200 \times 200 \times 200$, and the results are comparable to that in the lower dimension model; however, the calculation time is ~ 6 times greater than that. Therefore, to determine the reason, further optimization for the model is ongoing.

Transferability. To test the transferability of DeepNCI, the transfer learning for small-scale HBDE data has been performed. The descriptors of HBDE are shown in Table S7 in the Supporting Information. In the transferability test, the weights from trained DeepNCI are taken as the initialization of transfer learning, and the model is fine-tuned by the training set of HBDE data without changing hyperparameters and the framework. A comparison has been made on the prediction for the same dataset obtained in MIV-BPNN and GP-GRNN⁴⁹ to determine if transfer learning is feasible. Table 3 gives the RMSEs of HBDE data by different methods. It can be seen that the transfer learning result is similar to the highest precision among simple machine learning models. This indicates that DeepNCI can offer a good starting point for another related or similar task. For analyzing extracted features for HBDE, t-SNE visualization of CNN-abstacted features and merged features on HBDE data are displayed in Figure 8. Similarly, the fine-tuned model built on extracted features can aggregate intraclass or discriminate interclass HBDE data. Apparently, the rambling data in Figure 8a become an ordered distribution in Figure 8b, because of extracted important features by the fused model. From above transfer learning results, it suggests that DeepNCI can be transferred to relative or similar tasks, and, in this way, solving some small sampling problems becomes possible.

CONCLUSION

In this work, a novel multimodal deep architecture regression model, DeepNCI, is proposed to correct the DFT-calculated NCIs. Two networks dealing with DFT calculated electron density and quantum chemical properties are fused to build the multimodal regression. The multimodal can take effective information from multiple channels of its subnetworks; therefore, predictions with the multimodal architecture outperform either subnetwork. The DFT M062X/6-31G*

calculations by DeepNCI are improved to ~ 0.2 kcal/mol deviating from the CCSD(T)/CBS calculation. From the model performance and the t-SNE feature visualization, it can be found that DeepNCI can generate representative and comprehensive features well to distinguish molecules from different NCI dominant categories. A defined KNN AD can explicitly show the uncertainty of a prediction for user applications. Moreover, the model can be transferred to related target modeling, and thus, some small sampling problems may be solved via its transferability.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.1c01305>.

The quantum chemical descriptors, predictions for the test set and two external test sets, the hyperparameter setting and the training loss curve for DeepNCI, HDDBE database and additional results as well as figures ([PDF](#))

Accession Codes

The NCI database was downloaded from <http://www.begdb.org>. The source code of the DeepNCI model can be freely accessed at <https://github.com/wenzelee/DeepNCI>.

AUTHOR INFORMATION

Corresponding Authors

LiHong Hu – School of Information Science and Technology, Northeast Normal University, Changchun 130117, China; orcid.org/0000-0003-3792-2917; Email: lhhu@nenu.edu.cn

GuanHua Chen – Department of Chemistry, The University of Hong Kong, Hong Kong S.A.R., China; orcid.org/0000-0001-5015-0902; Email: ghc@everest.hku.hk

Authors

Wenze Li – School of Information Science and Technology, Northeast Normal University, Changchun 130117, China

Donghan Wang – School of Information Science and Technology, Northeast Normal University, Changchun 130117, China

Zirui Yang – School of Information Science and Technology, Northeast Normal University, Changchun 130117, China

Huijie Zhang – School of Information Science and Technology, Northeast Normal University, Changchun 130117, China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jcim.1c01305>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge for the financial support from NSFC (National Natural Science Foundation of China, No. 21473025 and the RGC General Research Fund (Grant No. 17309620) and Hong Kong Quantum AI Lab Limited.

REFERENCES

- (1) Mundlapati, V. R.; Sahoo, D. K.; Bhaumik, S.; Jena, S.; Chandrakar, A.; Biswal, H. S. Noncovalent carbon-bonding interactions in proteins. *Angew. Chem., Int. Ed.* **2018**, *57* (50), 16496–16500.

- (2) Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. Revealing Noncovalent Interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498–6506.
- (3) Zhan, H.; Liang, J. F. Extreme Activity of Drug Nanocrystals Coated with A Layer of Non-Covalent Polymers from Self-Assembled Boric Acid. *Sci. Rep.* **2016**, *6*, 38668.
- (4) Tiku, V.; Dikic, I. Autophagy without conjugation. *Nat. Struct. Mol. Biol.* **2019**, *26*, 249–250.
- (5) Biedermann, F.; Schneider, H.-J. Experimental Binding Energies in Supramolecular Complexes External Link. *Chem. Rev.* **2016**, *116*, 5216–5300.
- (6) Schulz, N.; Sokkar, P.; Engelage, E.; Schindler, S.; Erdelyi, M.; Sanchez-Garcia, E.; Huber, S. M. The interaction modes of haloimidazolium salts in solution. *Chem. - Eur. J.* **2018**, *24*, 3464–3473.
- (7) Lu, T.; Zhang, J.; Chen, J.; Gou, Q.; Xia, Z.; Feng, G. Structure and non-covalent interactions of 1,3-difluoropropane and its complex with water explored by rotational spectroscopy and quantum chemical calculations. *J. Chem. Phys.* **2019**, *150*, 064305.
- (8) Pan, H. A non-covalent dimer formed in electrospray ionisation mass spectrometry behaving as precursor for fragmentations. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3555–3560.
- (9) Mahadevi, A. S.; Sastry, G. N. Cooperativity in noncovalent interactions. *Chem. Rev.* **2016**, *116*, 2775–2825.
- (10) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- (11) *Theory and Applications of Computational Chemistry: The First Forty Years*, 1st Edition; Dykstra, C., Frenking, G., Kim, K., Scuseria, G., Eds.; Elsevier, 2005.
- (12) Tajti, A.; Szalay, P. G.; Csaszar, A. G.; Kallay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vazquez, J.; Stanton, J. F. HEAT: High accuracy extrapolated ab initio thermochemistry. *J. Chem. Phys.* **2004**, *121*, 11599.
- (13) Bursch, M.; Kunze, L.; Vibhute, A. M.; Hansen, A.; Sureshan, K. M.; Jones, P. G.; Grimme, S.; Werz, D. B. Quantification of Noncovalent Interactions in Azide–Pnictogen, –Chalcogen, and–Halogen Contacts. *Chem. - Eur. J.* **2021**, *27* (14), 4627–4639.
- (14) Zhang, I.; Xu, X. Exploring the Limits of the XYG3-Type Doubly Hybrid Approximations for the Main-Group Chemistry: The xDH@ B3LYP Model. *J. Phys. Chem. Lett.* **2021**, *12*, 2638–2644.
- (15) Schütt, K. T.; Gastegger, M.; Tkatchenko, A.; Müller, K.-R.; Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **2019**, *10*, 5024.
- (16) Cheng, L.; Welborn, M.; Christensen, A. S.; Miller, T. F. A universal density matrix functional from molecular orbital-based machine learning: transferability across organic molecules. *J. Chem. Phys.* **2019**, *150*, 131103.
- (17) Sugawara, M. Numerical solution of the Schrödinger equation by neural network and genetic algorithm. *Comput. Phys. Commun.* **2001**, *140*, 366–380.
- (18) Manzhos, S.; Carrington, T. An improved neural network method for solving the Schrödinger equation. *Can. J. Chem.* **2009**, *87*, 864–871.
- (19) Hu, L. H.; Wang, X. J.; Wong, L. H.; Chen, G. H. (2003) Combined first-principles calculation and neural network correction approach for heat of formation. *J. Chem. Phys.* **2003**, *119*, 11501–11507.
- (20) Gao, T.; Li, H.; Li, W.; Li, L.; Fang, C.; Li, H.; Hu, L.; Lu, Y.; Su, Z. A machine learning correction for DFT non-covalent interactions based on the S22, S66 and X40 benchmark databases. *J. Cheminform.* **2016**, *8*, 1–17.
- (21) Li, W.; Miao, W.; Cui, J.; Fang, C.; Su, S.; Li, H.; Hu, L.; Lu, Y.; Chen, G. Efficient Corrections for DFT Noncovalent Interactions Based on Ensemble Learning Models. *J. Chem. Inf. Model.* **2019**, *59*, 1849–1857.
- (22) Cui, J.; Li, W.; Fang, C.; Su, S.; Luan, J.; Gao, T.; Hu, L.; Lu, Y.; Chen, G. AdaBoost Ensemble Correction Models for TDDFT Calculated Absorption Energies. *IEEE Access* **2019**, *7*, 38397–38406.
- (23) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; Von Lilienfeld, O. A. Big data meets quantum chemistry approximations: The Δ -machine learning approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.
- (24) Wilkins, D. M.; Grisafi, A.; Yang, Y.; Lao, K. U.; DiStasio, R. A., Jr.; Ceriotti, M. Accurate molecular polarizabilities with coupled cluster theory and machine learning. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 3401–3406.
- (25) Gastegger, M.; Behler, J.; Marquetand, P. Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* **2017**, *8*, 6924–6935.
- (26) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; Von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
- (27) Mezei, P. D.; Von Lilienfeld, O. A. Noncovalent quantum machine learning corrections to density functionals. *J. Chem. Theory Comput.* **2020**, *16*, 2647–2653.
- (28) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, July 2017; pp 1263–1272.
- (29) Smith, J. S.; Isayev, O.; Roitberg, A. E. (2017) ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192–3203.
- (30) Li, H.; Zhong, Z. Y.; Li, L.; Gao, R.; Cui, J. X.; Gao, T.; Hu, L.; Lu, Y.; Su, Z. M.; Li, H. A cascaded QSAR model for efficient prediction of overall power conversion efficiency of all-organic dye-sensitized solar cells. *J. Comput. Chem.* **2015**, *36*, 1036–1046.
- (31) Mater, A.; Coote, M. Deep Learning in Chemistry. *J. Chem. Inf. Model.* **2019**, *59*, 2545–2559.
- (32) Lim, J.; Ryu, S.; Kim, J. W.; Kim, W. Y. Molecular generative model based on conditional variational autoencoder for de novo molecular design. *J. Cheminf.* **2018**, *10*, 1–9.
- (33) Ma, J.; Sheridan, R. P.; Liaw, A.; Dahl, G. E.; Svetnik, V. Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* **2015**, *55*, 263–274.
- (34) Kusumoto, D.; Seki, T.; Sawada, H.; Kunitomi, A.; Katsuki, T.; Kimura, M.; Ito, S.; Komuro, J.; Hashimoto, H.; Fukuda, K.; Yuasa, S. Anti-senescent drug screening by deep learning-based morphology senescence scoring. *Nat. Commun.* **2021**, *12*, 257.
- (35) Hermann, J.; Schätzle, Z.; Noé, F. Deep-neural-network solution of the electronic Schrödinger equation. *Nat. Chem.* **2020**, *12*, 891–897.
- (36) Lotter, W.; Diab, A. R.; Haslam, B.; Kim, J. C.; Grisot, G.; Wu, E.; Wu, K.; Onieva, J. O.; Boyer, Y.; Boxerman, J. L.; Wang, M.; Bandler, M.; Vijayaraghavan, G. R.; Sorensen, A. G. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat. Med.* **2021**, *27*, 244–249.
- (37) Yao, K.; Parkhill, J. The Kinetic energy of hydrocarbons as a function of electron density and convolutional neural networks. *J. Chem. Theory Comput.* **2016**, *12*, 1139–1147.
- (38) Segler, M. H. S.; Kogej, T.; Tyrchan, C.; Waller, M. P. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.* **2018**, *4*, 120–131.
- (39) Nigam, A. K.; Friederich, P.; Krenn, M.; Aspuru-Guzik, A. Augmenting Genetic Algorithms with Deep Neural Networks for Exploring the Chemical Space. In *International Conference on Learning Representations*, Addis Ababa, Ethiopia, April 25–30, 2020.
- (40) Rathi, P. C.; Ludlow, R. F.; Verdonk, M. L. Practical high-quality electrostatic potential surfaces for drug discovery using a graph-convolutional deep neural network. *J. Med. Chem.* **2020**, *63*, 8778–8790.
- (41) Torng, W.; Altman, R. B. Graph convolutional neural networks for predicting drug-target interactions. *J. Chem. Inf. Model.* **2019**, *59*, 4131–4149.

- (42) Zhou, Y.; Wu, J.; Chen, S. G.; Chen, G. H. Toward the exact exchange–correlation potential: A three-dimensional convolutional neural network construct. *J. Phys. Chem. Lett.* **2019**, *10* (22), 7264–7269.
- (43) Lim, J.; Ryu, S.; Park, K.; Choe, Y. J.; Ham, J.; Kim, W. Y. Predicting drug–target interaction using a novel graph neural network with 3D structure-embedded graph representation. *J. Chem. Inf. Model.* **2019**, *59*, 3981–3988.
- (44) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced database of benchmark Interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.
- (45) Gráfová, L.; Pitonák, M.; Řezáč, J.; Hobza, P. Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended S22 data set. *J. Chem. Theory Comput.* **2010**, *6*, 2365–2376.
- (46) Rezáč, J.; Riley, K. E.; Hobza, P. Benchmark calculations of noncovalent interactions of halogenated molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.
- (47) 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., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghuvaran, 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 D.01*; Gaussian, Inc., Wallingford, CT, 2013.
- (48) Li, W.; Fang, C.; Liu, J.; Cui, J.; Li, H.; Gao, T.; Li, H.; Hu, L.; Lu, Y. HSPXY: A hybrid-correlation and diversity-distances based data partition method. *J. Chemom.* **2019**, *33*, e3109.
- (49) Li, H. Improving the accuracy of density functional theory calculation for homolysis bond dissociation energies of Y-NO bond: Neural network and support vector machine methods (in Chin.). Ph.D. Dissertation, Northeast Normal University, 2010.
- (50) Rezac, J.; Hobza, P. Describing noncovalent interactions beyond the common approximations: How accurate is the "Gold Standard," CCSD(T) at the complete basis set limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.
- (51) Berka, K.; Laskowski, R.; Riley, K. E.; Hobza, P.; Vondrasek, J. Representative amino acid side chain interactions in protein. A comparison of highly accurate correlated ab initio quantum chemical and empirical procedures. *J. Chem. Theory Comput.* **2009**, *5*, 982–992.