Improving the Performance of Long-Range-Corrected Exchange-Correlation Functional with an Embedded Neural Network

Qin Liu, JingChun Wang, PengLi Du, LiHong Hu, Xiao Zheng, and GuanHua Chen

ABSTRACT: A machine-learning-based exchange-correlation functional is proposed for general-purpose density functional theory calculations. It is built upon the long-range-corrected Becke–Lee–Yang–Parr (LC–BLYP) functional, along with an embedded neural network which determines the value of the range-separation parameter \( \mu \) for every individual system. The structure and the weights of the neural network are optimized with a reference data set containing 368 highly accurate thermochemical and kinetic energies. The newly developed functional (LC–BLYP–NN) achieves a balanced performance for a variety of energetic properties investigated. It largely improves the accuracy of atomization energies and heats of formation on which the original LC–BLYP with a fixed \( \mu \) performs rather poorly. Meanwhile, it yields a similar or slightly compromised accuracy for ionization potentials, electron affinities, and reaction barriers, for which the original LC–BLYP works reasonably well. This work clearly highlights the potential usefulness of machine-learning techniques for improving density functional calculations.

1. INTRODUCTION

Machine learning (ML), a fundamental and essential approach to artificial intelligence, has prompted a wide range of applications. These include robotics,\(^4\)\(^-\)\(^7\) automatic processing and translation of languages,\(^8\)\(^-\)\(^9\) pattern recognition,\(^8\)\(^-\)\(^9\) drug design,\(^10\)\(^-\)\(^12\) and game playing.\(^13\) Machine learning has also been increasingly adopted by studies in various fields of physics, chemistry, and materials science.\(^14\)\(^-\)\(^17\)

In the field of theoretical chemistry, there have been many attempts to apply ML methods, particularly the neural network (NN) approach, to improve the accuracy of quantum-mechanical calculations/simulations on chemical systems. Being the most popular first-principles method, density functional theory (DFT) has been widely used in many fields of modern science. Despite their success, the present-day DFT methods have not achieved chemical accuracy for general cases. This is because the exact exchange-correlation (XC) functional is yet out of reach, while approximate functionals suffer from intrinsic errors.\(^18\)\(^-\)\(^19\) Tracing and eliminating the errors of density functional approximations is thus an important goal of theoretical advancement, and ML could be a useful tool for achieving this goal.

In 2003 Hu and Chen et al. developed a DFT–NEURON approach\(^20\) to improve the accuracy of heats of formation (HOF) of molecular species calculated by DFT methods. In their work, an artificial NN was constructed, whose weights were determined via a training process with a database containing 180 small- to medium-sized organic molecules. The optimized NN can then be used to correct the calculated HOF for new molecules. Such an approach was later generalized to improve density functional calculations for a variety of thermochemical properties, including Gibbs free energy, ionization potential (IP), electron affinity (EA), and absorption energy.\(^21\) In 2014, the DFT–NEURON approach was further improved by incorporating statistical methods such as Kennard–Stone sampling and bootstrapping methods.\(^22\) For a database containing 539 molecules, the NN correction reduces the mean absolute error (MAE) of HOF calculated by the hybrid Becke-3-Lee–Yang–Parr (B3LYP) functional\(^23\)\(^-\)\(^24\) from more than 10 kcal/mol to about 1 kcal/mol, and hence achieves the chemical accuracy.\(^25\) Recently, Hu and co-workers have established an NN to improve the description of noncovalent interactions by DFT methods.\(^25\) In 2007, Wu and Xu proposed the X1 method,\(^26\) which uses an NN to improve the prediction of thermochemical properties of molecular systems by DFT methods. The X1 method adopts an extensive data set comprised of diversified chemical species to train the NN, along with a multipopulation genetic algorithm to provide a reasonable initialization for the optimization of synaptic weights.\(^27\) Applications of X1 and its extensions have...
led to substantial alleviation of size-dependent errors in the calculated HOF, bond dissociation enthalpies, and heats of isomerization.²⁸

There have also been some efforts that aim to improve the approximate XC functionals with ML techniques. For instance, in 2004 some authors of the present work have constructed an NN to optimize the three parameters in B3LYP for every individual molecule, and the resulting ML-based B3LYP predicts more accurate thermochemical properties than the conventional B3LYP functional.²⁹ It is worth pointing out that, although the optimal parameter values are system-dependent, the form and weights of the NN are fixed for all systems, and thus the same NN can be universally applied to any molecular system. Recently, Snyder et al. have applied ML to determine the kinetic energy of a one-dimensional model electronic system, which achieves chemical accuracy.³⁰,³¹

Even more ambitious strategies to combine ML with DFT methods have been proposed. Very recently, Smith et al. have demonstrated that a deep NN trained on DFT calculations can learn an accurate and transferable potential for organic molecules. Brockherde et al. have used ML to directly learn the density-potential and energy-density maps,³²,³³ and thus bypass the need of solving Kohn–Sham equations.³⁴

Despite the encouraging progress, ML-based DFT methods are yet far from mature, as they have not been applied to realistic systems for general-purpose usage. For instance, although our previously developed ML-based B3LYP functional³⁵ has demonstrated a promising performance, it has a few notable limitations: (1) the training data set was rather small, (2) the quality of descriptors was not systematically evaluated, and (3) the three parameters were optimized simultaneously within the same NN, and their relative significance was not analyzed. In view of the exciting development of ML during the past decade, we are keen to revisit the ML-based XC functional and address the above issues.

In this work, we choose to apply ML to the long-range-corrected (LC) functional. The range-separation scheme was originally proposed by Savin et al.,³⁵ based on which Hirao and co-workers have developed the LC functionals.³⁶–³⁹ An LC functional includes a certain amount of long-range Hartree–Fock (HF) exchange, and thus gives rise to improved description of 4s–3d interconfigurational energies of the first-row transition metals,³⁹ polarizabilities of π-conjugated systems,³⁹ reaction barriers,⁴⁰ charge-transfer excitation energies and Rydberg excitation energies,³⁶,³⁸ and van der Waals and π-aromatic interactions.⁴¹,⁴² However, LC functionals are known to yield less accurate HOF.⁴⁰ Most LC functionals involve a range-separation parameter , which quantifies how the electron–electron interaction operator is decomposed into long-range and short-range parts (see Section 2 for details). The parameter has been set to different values for calculation of different properties. For instance, has been set to 0.33 au throughout this paper unless specified otherwise) to minimize the MAE of the internuclear distances of the first-to-third-row homonuclear diatomic molecules.³⁸ The value of has also been set to 0.47 au to minimize the error of calculated atomization energies (AEs).⁴⁶ Other values of 0.15,⁴⁷–⁴⁹ 0.4,⁴⁵ and 0.5 au⁵⁰–⁵² have also been reported. Clearly, optimizing for a specific property is somewhat against the universal nature of the XC functional.

We consider the parameter to be a certain functional of electron density and it is thus system-dependent. We construct an NN that determines the most reasonable value of for every individual species, and the NN-predicted is then adopted to calculate all the thermochemical and kinetic properties of that species. Compared with our previous effort, we now use a more sophisticated NN, along with a more extensive data set and a group of carefully selected descriptors to perform the training. Again, although the predicted varies from system to system, the structure and weights of the NN are fixed for all systems, and thus the same NN can be universally applied to any molecular system.

The remainder of this paper is organized as follows: Section 2 elaborates the methodology adopted for constructing the ML-based functional. Specifically, details about the NN approach used and related computational techniques will be introduced. Numerical results are presented and discussed in Section 3. Finally, the concluding remarks are given in Section 4.

2. METHODOLOGY

2.1. Long-Range Correlated Functional. In the range-separation scheme of XC functional, the electron–electron interaction operator is split into short-range and long-range components as follows:

\[ \frac{1}{r} = \frac{1 - \text{erf}(\mu r)}{r} + \frac{\text{erf}(\mu r)}{r} \]  

(1)

Here, \( r = |r_1 - r_2| \) is the distance between electrons 1 and 2, erf(\( x \)) is the error function, and \( \mu \) is the range parameter of our primary interest.

We choose to examine the LC–BLYP functional, for which the XC energy is expressed as
where $E^\text{LR-HF}_x$ is the long-range component of the HF exchange functional, $E^\text{SR-B88}_x$ is the short-range component of the Becke’s generalized gradient approximation for exchange functional, and $E^\text{LYP}_x$ is the Lee–Yang–Parr (LYP) correlation functional.\textsuperscript{24}

### 2.2. Machine-Learning Based Exchange-Correlation Functional

2.2.1. Overview of Neural Network Construction and Training. Figure 1 is a schematic diagram that illustrates the architecture of NN and the workflow of training process. The NN constructed in this work is a three-layer backpropagation NN,\textsuperscript{55} which consists of an input layer, a hidden layer, and an output layer. An unsupervised learning network, autoencoder,\textsuperscript{66,67} is introduced at the pretraining stage (see the lower-left part of Figure 1). The autoencoder transforms the descriptors (information representing an individual species) of molecules to certain quantitative features, which are then fed to the input layer of NN. The autoencoder is also used to select the most informative descriptors and thus reduce the dimension of NN. The right part of Figure 1 illustrates the processing of data set. The reference data set contains 368 accurate energetic data of 362 molecules (details to be presented later). We also establish a numerical database to store the total energies of all the molecules using the LC–BLYP functional with $\mu$ varying from 0.2 to 0.7 in units of a.u. LC–BLYP is the Lee–Yang–Parr (LYP) correlation functional.\textsuperscript{54} Here, $\mu$ values determined by GA results in a set of molecules that are mutually different but all accurately reproduce the energies in the reference data set. We save all these $\mu_0$ sets for the subsequent training of NN.

2.2.2. Reference Data Set and Numerical Database. For numerical convenience, we split the training of NN into two steps: (1) determine the most reasonable value of $\mu$ for each individual molecule related to the reference data set; and (2) take these system-specific $\mu$ values (the $\mu_0$ set) as the target of NN to determine the weights.

To accomplish the first step, we set up a database to store the calculated total energies of the 362 chemical species with the LC–BLYP functional. For each species, calculation is done for $\mu$ varying from 0.2 to 0.7 with an increment of 0.005, and the total energy at any other value of $\mu$ can be obtained via a linear interpolation based on the recorded data. The geometries of the 362 species are optimized at the B3LYP/6-311++g-(3df,3pd) level. Zero-point energies (ZPEs) are obtained through frequency analysis. The total energies and thermochemical properties are computed by single-point calculations at the LC-BLYP/6-311++g(3df,3pd) level.\textsuperscript{68} All the DFT calculations are carried out by using the Gaussian09 suite of program.\textsuperscript{69}

The reference data set of energetic data contains 368 atomization energies (AEs), 51 ionization potentials (IPs), and 26 electron affinities from the G2/97 set,\textsuperscript{61,62} 75 harmonic frequencies from the G3/99 set,\textsuperscript{64,65} and 36 hydrogen transfer reaction energy barriers (HTREBs) and 32 non-hydrogen transfer reaction energy barriers (NHTREBs) from the H00H38/04 and NHTBH38/04 set, respectively.\textsuperscript{65–67}

There are totally 362 atom and molecular species associated with the reference data set, including 228 neutral molecules, 51 cations, 29 anions, 34 transition-state complexes, 14 atoms, and 6 molecular complexes bound by weak interactions. These chemical species are comprised of non-metallic elements H, B, C, N, O, F, Si, P, S, Cl and metallic elements Be, Li, Na, Al. The diversity of these chemical species is important for preserving the universality of ML-based functional.

Numerical Database. For numerical convenience, we split the training of NN into two steps: (1) determine the most reasonable value of $\mu$ for each of the 362 species that are associated with the reference data set. This is done by employing the GA, which is a metaheuristic and adaptive algorithm based on the evolutionary ideas of natural selection.\textsuperscript{70,71} Note that a molecule may be associated with more than one energetic data, and hence the determination of $\mu_0$ is to reproduce accurately the reference energies data set in an average sense. Mathematically, the following criterion is adopted:

$$\mu_0 = \arg \min_{\mu} \sum_{i=1}^{368} |E^\text{DFT}_i(\mu) - E^\text{ref}_i|$$  

Here, $E^\text{ref}_i$ is the $i$th reference energy in the data set, and $E^\text{DFT}_i(\mu)$ is the same energy calculated using the LC–BLYP functional with the system-specific $\mu$ values. The GA results in a number of $\mu_0$ sets that are mutually different but all accurately reproduce the energies in the reference data set. We save all these $\mu_0$ sets for the subsequent training of NN.

Figure 2 depicts the distribution of system-specific $\mu$ values for a typical $\mu_0$ set determined by the GA. Interestingly, the $\mu$ values are roughly normally distributed, with about 68% of data falling within the range of (0.35,0.55) and about 80% within (0.3,0.6) (in units of a.u.). LC–BLYP calculations adopting...
such a $\mu_0$ set reproduce accurately the thermochemical and kinetic energies in the reference data set with an MAE of 0.31 kcal/mol, which is substantially reduced from the MAE of 12.24 kcal/mol by conventional LC–BLYP functional with $\mu = 0.47$ a.u. fixed for all chemical species.

2.2.3. Selection of Descriptors. Descriptors. In the context of NN, descriptors are essentially important as they uniquely define a sample in the data set. Since the goal of our present work is to find a mapping between the $\mu$-parameter and the electron density of a chemical species, the descriptors should characterize or represent to certain extent the electron density of the species.

Following previous studies, we consider three categories of descriptors: (1) The first category includes the basic physical quantities of a species, such as the number of atoms $N_m$ the number of electrons $N_e$, the spin multiplicity $g_s = 2S + 1$ with $S$ being the total spin, and number of atoms for a specific element ($N_{Li}$, $N_{Be}$, $N_{B}$, $N_{C}$, $N_{N}$, $N_{O}$, $N_{F}$, $N_{Ne}$, $N_{Na}$, $N_{Mg}$, $N_{Al}$, $N_{Si}$, $N_{P}$, $N_{S}$, $N_{Cl}$). (2) The second category concerns some characteristic properties of the species, such as the ZPE, the dipole moment (DM), the energy of highest occupied molecular orbital (HOMO), the energy of lowest unoccupied molecular orbital (LUMO), the energy gap $\epsilon_{gap}$ etc. These descriptors are obtained from DFT calculations at the LC–BLYP/6-311+g(3df,3pd) level with the preset $\mu = 0.47$ a.u. (3) The third category concerns the geometric structure of the molecule, represented by the PAE, the Pauling electronegativity $\eta$, and $\eta_{A}$ for the different atoms, and it is the Pauling electronegativity 7,78 of atom $i$. $W_{AA}$ characterizes the pairwise interactions between different atoms, and it conveys the structural information on the species (via $r_{ij}$), while $W_{AA}$ describes the self-interactions of atoms. Different from their original definitions in ref 3, here both $W_{AA}$ and $W_{AA}$ are averaged by $N_m$.

### Autoencoder

The above 24 descriptors are proposed by experience or by physical/chemical intuition. To simplify the NN model and to avoid possible overfitting, it is necessary to assess these descriptors and select the most informative and representative ones.

The autoencoder has become a more widely used pretraining tool to extract principal features from descriptors and to reduce the dimension of NN model. The autoencoder adopted in this work is sketched in Figure 3, which consists of the encoder and decoder.

![Figure 3. Structure and work flow of the autoencoder for the generation of features and selection of descriptors.](image)

The encoder transforms the descriptors $x$ at the input layer to the features $h$ at the hidden layer, and the decoder does an inverse transform from $h$ to $\hat{x}$ at the output layer, where $\hat{x}$ is an attempted replica of $x$. Mathematically, the encoding and decoding of descriptors are represented by the mappings $f$ and $g$ as follows.

$$ h = f(W_{x}x + b_{1}) $$

$$ \hat{x} = g(W_{h}h + b_{2}) $$

Here, $f$ and $g$ are realized by the logistic sigmoid function $f(x) = g(x) = 1/(1+e^{-x})$. $b_{1}$ and $b_{2}$ are the bias vectors introduced at the input and hidden layers, respectively. $W_{x}(W_{h})$ is the weight matrix between the input (hidden) layer and hidden (output) layer. Through an unsupervised training, the weight matrices are adjusted until the reconstruction error $L(x,\hat{x}) = ||x - \hat{x}||$ is minimized.

For each $(n)$th descriptor, we perform a linear regression between its values at the input ($x$) and the duplicated values at the output ($\hat{x}$) for all the $M = 362$ species associated with the reference data set. The coefficient of determination $R^2$ of the regression is computed as

$$ R^2 = 1 - \frac{\sum_{k=1}^{M}(x_{ik} - \bar{x}_{i})^2}{\sum_{k=1}^{M}(x_{ik} - \bar{x})^2} $$

where $\bar{x}_{i} = \frac{1}{M} \sum_{k=1}^{M} x_{ik}$ is the mean of $x_{i}$. Statistically, $R^2$ being close to 1 indicates that the information on descriptor is well preserved by the transformed features $h$. In contrast, a small $R^2$ implies loss of information through the transforms.

Among all the 24 descriptors defined above, there are 10 descriptors which have an $R^2$ value larger than 0.89, while the $R^2$ for the rests are all smaller than 0.62. The 10 highly reproducible descriptors ($N_{N}$, $N_{P}$, $N_{Ne}$, $N_{Cl}$, $N_{F}$, ZPE, DM, $W_{AA}$, and $W_{AA}$) are thus employed for the subsequent training of NN. We have manually selected various combinations of descriptors to train the NN. The results confirm that the above 10 descriptors yield the best training performance. All the numerical procedures related to the autoencoder and the NN are carried out with the Neural Network Toolbox of MATLAB.

2.2.4. Training of Neural Network. The structure of NN for predicting the system-specific $\mu$ values is detailed in Figure 4. The input, hidden and output layers contain 10, 5, and 1 neurons, respectively. Given the limited size of the training set, further increasing the number of hidden layers or hidden neurons would result in an excessive number of weight parameters, which could easily lead to overfitting. The features ($h$) of the chemical species related to the training data set are
generated by the autoencoder and enter the NN at the input layer. The output of the NN, the \( \mu \) value, is expressed as

\[
\mu = g(W'_1 f(W'_2 h + b'_1)) + b'_2
\]

(9)

Here, \( f(g) \) is the transfer function which connects the input (hidden) layer to the hidden (output) layer, with \( W'_1 \) (\( W'_2 \)) being the associated weight matrix and \( b'_1 \) (\( b'_2 \)) the bias vector. We choose the nonlinear tan-sigmoid function as the neural transfer function, i.e., \( f(x) = g(x) = 2/(1+e^{-2x}) - 1 \). The use of a linear transfer function yields much more inferior training performance. The goal of training is to minimize the mean squared deviation between \( \mu \) of eq 9 and the reference \( \mu_0 \) set by tuning the weights. Note that the value of \( \mu \) obtained through eq 9 ranges from \(-1\) to \(1\), and hence the \( \mu_0 \) set is scaled linearly to the same range for making the comparison.

The Bayesian regularization method is adopted to perform the training. For each training process, the data set is randomly divided into the training (85%) and testing (15%) subsets. The testing subset ensures that the final NN can be generalized to new species outside the reference data set.

The training of NN has been carried out numerous times with different initial weights and biases, different numbers of neurons, different combinations of descriptors, and different reference \( \mu_0 \) sets. The minimal mean squared deviation between \( \mu \) and \( \mu_0 \) achieved is about 0.004 au, which is considered to be a reasonably small value.

3. RESULTS AND DISCUSSION

We first examined how well the LC–BLYP functional with the NN-produced \( \mu \) values could reproduce the thermochemical and kinetic energies in the reference data set. As has been mentioned in Section 2.2.2, the LC–BLYP functional with the GA-generated \( \mu_0 \) set recovers accurately the energetic data with a rather small MAE of 0.31 kcal/mol. Even with the optimal training, the NN-produced \( \mu \) still differs from the \( \mu_0 \) set with a mean squared deviation of 0.004 au. Such a deviation is responsible for the errors in the calculated energies.

Figure 5. Distribution of the unsigned errors \( \Delta E \), for the 368 energetic data in the reference data set. The increment of error is 5 kcal/mol.
The Journal of Physical Chemistry A

errors (Δfi ≡ EDFT − Efi) for the original LC−BLYP and the LC−BLYP−NN. The original LC−BLYP gives rise to a rather large spread of errors, with a long tail extending into the negative error regime. The maximal negative error is −61.58 kcal/mol (the HOF of C6F6). In contrast, with the NN-produced μ values, the error distribution becomes roughly normal and much narrower. With the original LC−BLYP 56% of the calculated data have errors smaller than 10 kcal/mol and 42% are within the range of 5 kcal/mol; while with the LC−BLYP−NN the corresponding ratios increase substantially to 82% and 56%, respectively. Therefore, it is apparent that the application of NN indeed enhances the performance of the LC−BLYP in the statistical sense.

For practical usage, it is crucial that the LC−BLYP−NN functional is capable of producing the same level of accuracy for chemical species beyond the reference data set. To verify this point, we adopt a “prediction” data set that has no overlap with the reference data set. The prediction data set contains 13 HOF from the G3/05 set, 19 IPs and 16 EAs associated with the auxiliary neutral subset of the G2/97 set, and the heats of reaction (HOR) for 14 Diels–Alder reactions given in Figure 1 of ref. 85. The prediction data set involves 106 chemical species (including 71 neutral molecules, 19 cations, and 16 anions) which have no overlap with the species associated with the reference data set.

We then employ the established LC−BLYP−NN functional to reproduce the energetic data in the prediction data set. Numerical calculations proceed as follows: (1) Evaluate the 10 descriptors (Nν, Nσ, Nνb, Nσ, Nα, Nβ, ZPE, DM, Wc, and Wαc) of the species associated with the prediction data set. (2) Feed the descriptors to the optimized autoencoder and generate the features of the molecular species. (3) Import the features to the trained NN, which outputs the system-specific μ values. (4) Carry out LC−BLYP calculations for the thermochemical energies with the system-specific μ values.

Table 2 lists the MAEs of the calculated energetic data in the prediction data set. Evidently, the errors of HOF, IP, and EA

<table>
<thead>
<tr>
<th>Energy</th>
<th>Number</th>
<th>B3LYP</th>
<th>LC−BLYP</th>
<th>LC−BLYP−NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOF</td>
<td>13</td>
<td>7.06</td>
<td>21.71</td>
<td>10.68</td>
</tr>
<tr>
<td>IP</td>
<td>19</td>
<td>4.15</td>
<td>3.27</td>
<td>4.02</td>
</tr>
<tr>
<td>EA</td>
<td>16</td>
<td>2.51</td>
<td>3.10</td>
<td>3.49</td>
</tr>
<tr>
<td>HOR</td>
<td>14</td>
<td>13.71</td>
<td>1.13</td>
<td>6.52</td>
</tr>
<tr>
<td>Overall</td>
<td>62</td>
<td>6.50</td>
<td>6.61</td>
<td>5.85</td>
</tr>
</tbody>
</table>

*DFT results obtained from the B3LYP, the original LC−BLYP with a fixed μ = 0.47a.u., and the LC−BLYP with the NN-produced system-specific μ values are presented.

resulted from the original LC−BLYP and the LC−BLYP−NN are overall similar to those in the reference data set: the system-specific μ values largely improve the HOF while preserving roughly the same level of accuracy for IP and EA. Therefore, the errors are more balanced among the various types of energies with the use of NN. This is clearly verified by the error distribution shown in Figure 6. In particular, the original LC−BLYP much too low HOF for the molecules C6F6 and C6F5Cl with rather large unsigned errors of −79.97 kcal/mol and −69.47 kcal/mol, respectively; while with the LC−BLYP−NN the above errors are reduced to −4.58 kcal/mol and −6.16 kcal/mol, respectively.

Figure 6. Distribution of the unsigned errors Δfi for the 62 energetic data in the prediction data set. The increment of error is 5 kcal/mol.

We include a new type of thermochemical energies in the prediction data set, the HOR for the formation of a group of cyclic and cagelike molecules. Such molecules are distinctly different from the species related to the reference data set. The original LC−BLYP yields a remarkably high accuracy for the HOR with a very small MAE of 1.13 kcal/mol, while the LC−BLYP−NN results in a somewhat compromised accuracy with an MAE of 6.52 kcal/mol. Nevertheless, the latter MAE is still much smaller than that with the B3LYP functional.

The consistent behavior of the LC−BLYP−NN on the training and prediction data sets affirms that the applicability of the NN constructed in this work can be robustly extended to systems beyond the training data set, which is important not only for ensuring the predictive power of the NN, but also for preserving the universality of the XC functional.

Through our extensive numerical tests, it is found that many aspects of the NN, such as the size and diversity of the reference data set, the representability of descriptors, and the efficacy of pretraining and training algorithms, affect significantly the performance of the resulting functional. The present work excels our previous effort in all these aspects. Nevertheless, it is apparent that there is still plenty of room for improvement, considering the fact that the GA-generated μ set yields an MAE as small as 0.31 kcal/mol for the reference data set.

4. CONCLUDING REMARKS

To summarize, in this work we propose an ML-based XC functional, the LC−BLYP−NN functional, for general-purpose DFT calculations. The ML-based functional is built upon the explicit form of the conventional LC−BLYP functional, but with a system-specific range-separation parameter μ which is determined by a carefully designed and well-trained NN.

The newly constructed ML-based functional is tested extensively on a variety of thermochemical and kinetic energies. The LC−BLYP−NN achieves a balanced performance for the various types of properties investigated. It largely improves the accuracy of AE and HOF on which the original LC−BLYP with a fixed μ performs rather poorly. Meanwhile, it yields a similar or slightly compromised accuracy for IP, EA and reaction barriers, for which the original LC−BLYP works reasonably well.
Usually, the inaccurate or occasionally the problematic results of DFT calculations can be traced to the intrinsic errors associated with the approximate forms of XC functional. For instance, it is well-known that the local density approximation or the generalized gradient approximations (GGAs) suffer from the delocalization error,85–88 and hence they tend to predict too low energies for delocalized electron distributions.8,18,19 In contrast, the HF exchange functional is size-consistency condition, since 

\[ \mu \] 

agrees with the 

\[ \mu \] 

of the involving subsystems. Moreover, mainstream approximate functionals have many sources of errors. Besides the delocalization and localization errors mentioned above, density functional approximations also suffer from the self-interaction error,82,93 and the dynamic and static correlation errors.82 Systematic removal or alleviation of these errors is extremely challenging, which would require a functional form that is more sophisticated than the range-separation scheme. For instance, the LC−BLYP functional underestimates (although much less severely than the LDA and GGAs) the energies of \( H_2 \) and \( He_2 \) molecules at large internuclear distances,100 because the delocalization and localization errors associated with the short-range and long-range exchange parts do not cancel completely. The LC−BLYP−NN developed in this work does not improve this situation, since the presently adopted training set does not include energetic data explicitly representing the delocalization/ localization error (such as the deviation between integer and derivative gaps, or the energy of systems with fractional electron numbers).89 This will be a subject of our future work. In recent years, a variety of new functional forms have been proposed, such as the local hybrid hyper-GGA functionals,95 doubly hybrid functionals,96,98 and scaling corrected functionals.99 Combining the ML approaches with these new functional forms is expected to lead to a major boost in the performance of DFT methods. Work along this direction is in progress.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpca.7b07045.

Details on the training and prediction of neural network; optimized weight matrices and bias vectors (PDF)

### AUTHOR INFORMATION

**Corresponding Author**

*E-mail: xx58@ustc.edu.cn.*

**ORCID**

Xiao Zheng: 0000-0002-9804-1833

### REFERENCES


### ACKNOWLEDGMENTS

The support from the Ministry of Science and Technology (Grant Nos. 2016YFA0400900 and 2016YFA0200600), the National Natural Science Foundation of China (Grant Nos. 21573202 and 21233007), the Fundamental Research Funds for the Central Universities (Grant No. 2340000074), and the SuperComputing Center of USTC is gratefully acknowledged.

### Notes

The authors declare no competing financial interest.


