Nuclear binding energies in artificial neural networks

Lin-Xin[g](https://orcid.org/0000-0002-5626-0704) Zeng \bullet [,](https://orcid.org/0000-0002-6654-3644) ¹ Yu-Ying Yin \bullet , ¹ Xiao-Xu Dong \bullet , ¹ and Li-Sheng Geng \bullet ^{1,2,3,4,*}

¹*School of Physics, Beihang University, Beijing 102206, China*

²*Peng Huanwu Collaborative Center for Research and Education, Beihang University, Beijing 100191, China*

³*Beijing Key Laboratory of Advanced Nuclear Materials and Physics, Beihang University, Beijing 102206, China*

⁴*Southern Center for Nuclear-Science Theory (SCNT), Institute of Modern Physics, Chinese Academy of Sciences, Huizhou 516000, China*

(Received 10 October 2022; revised 11 November 2023; accepted 20 February 2024; published 25 March 2024)

The binding energy or mass is one of the most fundamental properties of an atomic nucleus. Precise binding energies are vital inputs for many nuclear physics and nuclear astrophysics studies. However, due to the complexity of atomic nuclei and the nonperturbative strong interaction, up to now, no conventional physical model can describe nuclear binding energies with a precision below 0.1 MeV, the accuracy needed by nuclear astrophysical studies. In this work, artificial neural networks (ANNs), the so-called "universal approximators", are used to calculate nuclear binding energies. We show that the ANN can describe all the nuclei in AME2020 with a root-mean-square deviation (RMSD) around 0.2 MeV, better than the best macroscopic-microscopic models, such as FRDM and WS4. The success of the ANN is mainly due to the proper and essential input features we identify, which contain the most relevant physical information, i.e., shell, paring, and isospin-asymmetry effects. We show that the well-trained ANN has excellent extrapolation ability and can predict binding energies for those nuclei inaccessible experimentally. In particular, we highlight the important role of "feature engineering" for physical systems where data are relatively scarce, such as nuclear binding energies.

DOI: [10.1103/PhysRevC.109.034318](https://doi.org/10.1103/PhysRevC.109.034318)

I. INTRODUCTION

The atomic nucleus is a quantum many-body system with an extremely complex structure [\[1\]](#page-7-0). As one of the most fundamental properties of atomic nuclei, binding energies (BE) can provide crucial information on nuclear shapes [\[2\]](#page-7-0), shell effects $[3,4]$, pairing effects $[5]$, and the disappearance as well as the emergence of magic numbers $[4,6]$. In addition, binding energies are essential inputs for superheavy nuclei syntheses [\[7\]](#page-7-0) and nuclear astrophysical studies [\[8\]](#page-7-0), e.g., the *r*-process [\[9,10\]](#page-7-0), x-ray bursts [\[11\]](#page-7-0), etc. Therefore, reliable theoretical predictions and experimental measurements of nuclear binding energies have always been at the frontier of nuclear physics [\[12–](#page-7-0)[14\]](#page-8-0).

In the latest atomic mass evaluation (AME 2020) [\[15\]](#page-8-0), the masses of 3556 nuclei (including measured and extrapolated) are compiled. However, various theoretical models predict that about 8000 to 10000 nuclei may exist [\[12](#page-7-0)[,13,16,17\]](#page-8-0), including most of those relevant in nuclear elements syntheses. Therefore, reliable and accurate theoretical predictions are urgently needed. Some of the most widely used theoretical models include the Weizsäcker-Skyrme (WS) model [\[18–22\]](#page-8-0), the relativistic mean field model (RMF) [\[17,23\]](#page-8-0), the Duflo-Zuker model (DZ) [\[24\]](#page-8-0), the Hartree-Fock-Bogoliubov model $[25–28]$, the finite-range droplet model (FRDM) $[12,29,30]$ $[12,29,30]$, and the RCHB [\[13\]](#page-8-0) and DRHBc [\[31\]](#page-8-0) models. Most of these models can describe the experimental data with a root-meansquare deviation (RMSD) ranging from about 0.3 MeV to several MeV. Among them, FRDM2012 [\[30\]](#page-8-0) achieved an RMSD of 0.570 MeV, while the Weizsäcker-Skyrme (WS4) [\[22\]](#page-8-0) model gives the best description with an RMSD of 0.298 MeV. In general, the macro-micro models, rather than the more "physical" microscopic models, perform better in describing nuclear masses because their parameters are determined by fitting all the (then available) experimental data.

In recent years, artificial neural networks (ANNs), as one of the most powerful machine learning methods, have been successfully applied in nuclear physics studies [\[9](#page-7-0)[,32\]](#page-8-0), e.g., binding energies [\[33–38\]](#page-8-0), charge radii [\[39–42\]](#page-8-0), α -decay halflives $[43]$, β-decay half-lives $[44]$, and fission fragment yields [\[45–47\]](#page-8-0).

The studies of nuclear binding energies (masses) can be divided into two categories, i.e., either fitting to the experimental data directly or to the residuals between experimental data and model predictions. In Refs. [\[34,38,48–51\]](#page-8-0), mass residuals are utilized to refine the theoretical models. In Refs. [\[34–36,38\]](#page-8-0), Bayesian neural networks are found to be able to describe nuclear binding energies with an RMSD ranging from 0.266 to 0.850 MeV. The RMSD obtained in the WS4 supplemented with light gradient boosting machine (LightGBM) is 0.170 ± 0.011 MeV [\[49\]](#page-8-0). The Bayesian machine learning (BML) method proposed in Ref. [\[52\]](#page-8-0) achieves an RMSD of 84 keV, the first crossing the 100 keV threshold.

However, fewer works study the experimental data directly. In Refs. [\[37,53\]](#page-8-0), feed-forward neural networks with different structures are explored. Reference [\[37\]](#page-8-0) yielded an RMSD of 1.84 MeV for 1071 nuclei contained in AME2016 [\[54\]](#page-8-0) as

^{*}lisheng.geng@buaa.edu.cn

FIG. 1. The architecture of a neural network consists of two input features, two hidden layers of nine and seven nodes, and one output layer.

the test set.¹ Reference [\[53\]](#page-8-0) applied the data augmentation technique to expand the data set. The RMSD decreased to 1.322 MeV for the test set within the training data region and 1.495 MeV for the new nuclei beyond the training data region. In Refs. [\[55,56\]](#page-8-0), mixed density networks with 12 physically motivated features [\[56\]](#page-8-0) or eight features constrained by the GK relation [\[55\]](#page-8-0) are devised to describe nuclear mass excesses. In the latter work [\[55\]](#page-8-0), an RMSD of 0.316 MeV for the test set and 0.186 MeV for the training set for nuclei with $Z \geqslant 20$ was achieved, whose performance is comparable to that of WS4 $[22]$.

In this work, we develop an ANN with seven input features of the most relevance. We find that among the 12 features studied in Ref. [\[56\]](#page-8-0), only six of them are effective in our network. Meanwhile, we find that taking GeLU [\[57\]](#page-8-0) as the activation function enhances the predictive power of the ANN. Our ANN provides a better description of nuclear binding energies than all the conventional models and shows good extrapolation ability.

This article is organized as follows. In Sec. II, we explain how to construct the ANN and determine the physically motivated input features. Results and discussions are presented in Sec. [III.](#page-2-0) A summary and outlook are provided in Sec. [IV.](#page-7-0)

II. THEORETICAL FORMALISM

In this section, we introduce the ANN and mass data we used in detail.

A. Artificial neural network

Generally speaking, an ANN is a supervised machine learning method regarded as a "universal approximator". The ANN used in this work is a fully connected feed-forward neural network consisting of one input layer with seven features, two hidden layers, and one output layer, as shown in Fig. 1. The inputs I_i and outputs O_i of layer *j* are connected as

follows:

$$
O_j = f(W_j \cdot I_j + b_j), \tag{1}
$$

where j runs over the input layer and the hidden layers, W_j are the weights, b_i are the bias, and f is the activation function to be specified. For the output layer, no activation function is needed.

Although, in principle, one could improve the description of BEs with either more hidden layers or more nodes in each hidden layer, one often ends up with the overfitting problem. By trial and error, we find that with two hidden layers and about 800 parameters, our ANNs can well describe the binding energies. For an ANN with *I* inputs, two hidden layers, and one output, denoted as [*I*, *H*1, *H*2, *O*], the number of parameters is $(I + 1) \times H1 + (H1 + 1) \times H2 +$ $(H2 + 1) \times O$. Table [I](#page-2-0) lists the nodes and number of parameters of the different ANNs investigated in the present work. Note that to understand better how the different input features affect the performance of ANNs, in addition to the default ANN with seven features, we also study three other ANNs, where two, four, and six features are used. For the activation function, we choose GeLU [\[57\]](#page-8-0), which performs better than Tanh. For the loss function, we use the standard mean absolute error (MAE)

$$
LOSS = \frac{\sum_{i=1}^{N} |BE_i^{\text{th}} - BE_i^{\text{exp}}|}{N}.
$$
 (2)

For numerical implementation, we use the optimized tensor library PYTORCH [\[58\]](#page-8-0) and employ the Adam algorithm [\[59\]](#page-8-0) with a learning rate 0.0001 and the decay constants 0.9 and 0.999. The weight matrices of our ANNs are initialized in PYTORCH with the same random seed.

A supervised ANN maps inputs to the desired outputs. In the present case, the output is the binding energy of a nucleus. An atomic nucleus is determined by its proton and neutron numbers, so one can naively take them as the only input. Nevertheless, it is well known that for small data sets, engineered features (in addition to these "fundamental features"), which encode essential information (priors) about the system under investigation, can play an invaluable role in enhancing the capacity of ANNs. Such a technique is widely used in nuclear physics studies (e.g., [\[39,40,55,56\]](#page-8-0)). In Ref. [\[39\]](#page-8-0), it was shown that in addition to *N* and *Z*, with two more features accounting for the pairing and shell-closure effects, one can describe the nuclear charge radii much better than the Bayesian models without these two features. In particular, one can describe the strong odd-even staggerings of the charge radii of the calcium and potassium isotopes. In Ref. [\[40\]](#page-8-0), it was shown that the description can be further improved with two more features accounting for isospin dependence and local anomalies.

In the studies of binding energies, besides the abovementioned pairing, shell-closure, and isospin dependence effects, many other features have been studied [\[43,56\]](#page-8-0). In the present work, the most relevant features are those just mentioned, i.e., pairing, shell-closure, and isospin dependence. The pairing effects are encoded in Z_{EO} and N_{EO} , which is 1 when *Z*/*N* is odd or 0 otherwise. Shell effects are introduced

¹The rather poor performance may be attributed to the fact that the MLP model has only been trained 800 epochs.

Model	Structure	Number of parameters	Input features
ANN ₂	[2, 35, 19, 1]	809	Z, N
ANN4	[4, 35, 17, 1]	805	Z, N, Z_{EO}, N_{EO}
ANN ₆	[6, 32, 17, 1]	803	$Z, N, Z_{EQ}, N_{EQ}, \Delta Z, \Delta N$
ANN7	[7, 32, 16, 1]	801	Z, N, Z_{EQ} , N _{EQ} , ΔZ , ΔN , ASY

TABLE I. Structure, number of parameters, and input features of the different ANNs studied in the present work.

via ΔZ and ΔN , the differences between *Z* and *N*, and the closest magic numbers. In this work, the magic numbers are 8, 20, 28, 50, 82, 126, and 184. As one moves away from the β -stability line, isospin-asymmetry becomes large. Therefore, we take into account this effect by introducing the seventh feature, ASY, which is defined as

$$
ASY = \left(1 - \frac{\kappa}{A^{1/3}} + \xi \frac{2 - |I|}{2 + |I|A}\right) I^2 A f_s,\tag{3}
$$

where the parameters κ , ξ , and f_s are taken from WS4 [\[22\]](#page-8-0).

B. Mass data

AME2020 [\[15\]](#page-8-0), in which the masses of 3556 nuclei (including measured and extrapolated ones) are compiled, is referred to as the data set in this work. We stress that one of our primary purposes is to provide reliable predictions that may be relevant for future applications, such as nuclear astrophysical studies. Therefore, it is necessary to use all the existing data for training. However, to test our model and to avoid overfitting, we follow the strategy adopted in Ref. [\[40\]](#page-8-0) to divide the data according to the time when they were reported, i.e., we use the masses compiled in AME2016 [\[54\]](#page-8-0) as the training set and those measured between 2016 and 2020 and compiled in AME2020 as the test set (test20). Based on this selection, there are 3434 nuclei in the training set and 122 in the test set.

III. RESULTS AND DISCUSSIONS

To quantify how well the ANN can describe nuclear binding energies in the training and test sets, we use the standard root-mean-square deviation (RMSD), $\sigma_{\rm rms}$, defined as

$$
\sigma_{\rm rms} = \sqrt{\sum_{i}^{N} \frac{\left(B E_{i}^{\rm th} - B E_{i}^{\rm exp}\right)^{2}}{N}},\tag{4}
$$

where BE_i^{th} are the ANN predictions, and BE_i^{exp} are the experimental binding energies contained in the training and test sets.² The $\sigma_{\rm rms}$ obtained with different numbers of input features are shown in Table II. For the entire set, the RMSD reduces from 1.18 MeV in ANN2 to 0.20 MeV in ANN7. This demonstrates unambiguously that engineered features that explicitly encode the pairing, shell, and isospin-asymmetry effects can significantly improve the capacity of ANNs to describe/predict nuclear binding energies. We stress that the

total number of parameters is similar for all four network structures. From ANN2 to ANN4, with the features (*Zeo* and N_{EO}), the descriptions improve not only the training set but also the test set. The RMSD decreases by almost 50 percent from ANN2 to ANN4. The explicit consideration of shell effects $(\Delta Z$ and $\Delta N)$ further improves the descriptions, and the RMSD for the entire set crosses the 0.3 MeV threshold. Finally, the explicit consideration of the asymmetry effects further improves the description, and the RMSD of ANN7 (for the training set and the entire set) falls below 0.2 MeV.

To put the performance of ANN7 into better perspective, we compare it with one of the most refined conventional models, WS4, which only studied those nuclei with $Z \ge 8$ and $N \ge 8$, i.e., only 3336 nuclei and 120 nuclei among those contained in our training and test sets. The corresponding $\sigma_{\rm rms}$'s are given in Table [III.](#page-3-0) Three things are noteworthy. First, ANN7 performs better than WS4 for all three sets of data. Second, removing the light nuclei with $Z < 8$ or $N < 8$, the RMSD of ANN7 decreases from 0.2 to 0.16 for the entire set. Third, from the training set to the test set, the RMSDs of ANN7 and WS4 increase. Somehow, surprisingly, in terms of percentage, the increase of WS4 is even more significant than that of ANN7. We note in passing that for the 2353 nuclei studied in WS4 [\[22\]](#page-8-0), ANN7 gives an RMSD of 0.15 MeV, which should be compared with that of WS4, 0.29 MeV.

Figures [2](#page-3-0) and [3](#page-3-0) provide more details on the deviations of the ANN predictions from the experimental data. As seen from Fig. [2,](#page-3-0) ANN2 performs relatively worse for even-even nuclei than their neighboring nuclei (even-odd or odd-odd). With two more features, Z_{EO} and N_{EO} , which take into account explicitly the pairing effects, ANN4 improves the description of even-even nuclei, and the $\sigma_{\rm rms}$ is reduced from 1.05 MeV to 0.63 MeV for the test set and from 1.18 MeV to 0.55 MeV for the training set. However, ANN4 does not capture the shell effects. One can see from the bottom panel of Fig. [2](#page-3-0) that the deviations between the ANN predictions and the experimental data are larger for those nuclei, with either proton or neutron

TABLE II. RMSDs for the training set (consisting of 3434 nuclei) and test set (consisting of 122 nuclei) achieved using different network structures.

		$\sigma_{\rm rms}$ (MeV)	
Model	Training set	Test set	Entire set
ANN ₂	1.18	1.05	1.18
ANN ₄	0.55	0.63	0.55
ANN ₆	0.29	0.51	0.30
ANN7	0.19	0.34	0.20

 2 In the present work, we do not distinguish between the measured and extrapolated masses compiled in the mass evaluations (AME2016 and AME2020).

TABLE III. Comparisons between ANN7 and the WS4 model [\[22\]](#page-8-0), for nuclei with $Z \ge 8$ and $N \ge 8$ compiled in AME2020, i.e., 3336 nuclei and 120 nuclei contained in our training set and test set, respectively.

	$\sigma_{\rm rms}$ (MeV)		
Model	Training set	Test set	Entire set
ANN7	0.15	0.34	0.16
WS4	0.42	1.30	0.47

number being magic. For doubly magic nuclei, the deviation is substantial. With the shell effects taken into account, the ANN6 successfully describes these nuclei, as seen from the upper panel of Fig. 3. A closer examination of Fig. 3 reveals that for heavy nuclei with $N \sim 175$ and light nuclei with $N \sim 20$ and $Z \sim 25$, the deviations are relatively large. We note that these nuclei are more neutron rich, and are newly compiled in AME2020. As a result, one can anticipate that an improved description can be achieved by considering a feature that explicitly considers isospin asymmetry. This is indeed the case. The $\sigma_{\rm rms}$ of ANN7 for the test set decreases from 0.51

FIG. 2. Absolute deviations of the ANN2 and ANN4 predictions from the experimental binding energies. The gray lines denote the magic numbers.

FIG. 3. Absolute deviations of the ANN6 and ANN7 predictions from the experimental binding energies. The gray lines denote the magic numbers.

MeV to 0.34 MeV, even though the deviations for some light nuclei are still relatively large.

It is instructive to examine how the different ANNs perform for nuclei in different mass regions. In Fig. [4,](#page-4-0) we show the distribution of $\sigma_{\rm rms}$ over different nuclei from light to heavy. It is clear that, on average, ANN7 performs the best, but for light nuclei with $A < 40$, ANN4 is the best. For nuclei with $120 < A < 160$, ANN7 and ANN6 work similarly well, but ANN6 is slightly better. The fact that ANN7 is better than ANN6 for light and heavy nuclei indicates that the ASY feature plays an important role. Even though for nuclei with $80 \leq A < 200$, their RMSDs are both small, as shown in Fig. [5,](#page-4-0) the generalizability of ANN7 is more robust.

It is interesting to study the performance of different network structures as one moves away from the β stability line. In Fig. [5,](#page-4-0) we decompose those nuclei in the test set into seven groups in terms of $|N - Z|$ to judge the predictive powers of ANNs in different isospin-asymmetry regions. ANN7 achieves the most stable and accurate predictions. For nuclei in the $|N - Z| > 30$ regions, the RMSDs between the predictions of ANN7 and the experimental data stay about only 200 keV.

Single and two-neutron separation energies are observables better suited to showcase the details of theoretical models,

FIG. 4. Distributions of $\sigma_{\rm rms}$ between every 40 mass numbers in test20.

particularly whether the shell closure and pairing effects are properly considered. They could be deduced from the binding energies as follows:

$$
S_n(Z, N) = BE(Z, N) - BE(Z, N - 1),
$$

\n
$$
S_{2n}(Z, N) = BE(Z, N) - BE(Z, N - 2).
$$
 (5)

In Figs. [6](#page-5-0) and [7,](#page-6-0) we compare the experimental oneand two-neutron separation energies with the predictions of four ANNs for the Ca, Ni, Sn, and Pb isotopic chains. For *Sn*, ANN2 cannot describe at all the odd-even staggerings, while ANN4 largely improves the situation. However, as is also reflected in Fig. [7,](#page-6-0) for nuclei close to the shell closures, the deviations are larger. ANN6 and ANN7, on the other

FIG. 5. σ_{rms} as functions of $|N - Z|$, which reflects the ability of ANNs to describe nuclei with large isospin asymmetries. Note that there is no nucleus in test20 with $50 < |N - Z| < 60$.

FIG. 6. Experimental single neutron separation energies in comparison with the ANN predictions.

hand, can describe all the nuclei, including the neutron-rich ones.

The predictions of a sound theoretical model should center around its mean value with small spreads. To check how the four ANNs perform in this perspective, we show in Fig. [8,](#page-6-0) the number of nuclei for which the deviations between theory and experiment fall between certain ranges. We further fit these counts with normal distributions. The mean of the Gaussian fit indicates the accuracy of the predictions, and the variance reflects the range of deviations. From Fig. [8,](#page-6-0) it can be seen that the Gaussian fits of ANN2 and ANN4 are pretty flat, and there are still many deviations over 0.7 MeV. In contrast, both ANN6 and ANN7 have very narrow distributions. In addition, most predictions by ANN7 deviate from their experimental counterparts by less than 0.7 MeV. In this sense, ANN7 not only predicts well but also is more certain.

Comparison with some recent works

In the past, most machine learning studies of nuclear binding energies adopted the residual approach, which fits the residuals between experimental data and the predictions of

an underlying theoretical model [\[34,38,48–51\]](#page-8-0). In the past two years, several studies fitting directly to binding energies appeared. In the following, we compare our study with two recent works.

The data augmentation technique, i.e., Gaussian noise augmentation, was found in Ref. [\[53\]](#page-8-0) to improve the predictions of ANNs. The number of nuclei in the training set expands from 1685 to 10110. The improvements in different MLPs from the perspective of RMSDs are from 18.86% to 30.50% in the test set in the training data region and from 23.47% to 36.33% in the test set beyond the training data region. We also tried to apply the data augmentation technique to our model but found that this technique barely affected our results.

In the latest work of Mumpower *et al.* [\[55\]](#page-8-0), the GK relation [\[60\]](#page-8-0) was added into the MDN as a soft constraint. For nuclei with $N \geqslant Z$, the GK relation reads

$$
BE(Z, N) \approx BE(Z - 2, N + 2) + BE(Z - 1, N)
$$

-
$$
BE(Z - 2, N + 1) + BE(Z, N + 1)
$$

-
$$
BE(Z - 1, N + 2),
$$
 (6)

FIG. 7. Experimental two-neutron separation energies in comparison with the ANN predictions.

while for nuclei with $N \le Z$, it reads

$$
BE(Z, N) \approx BE(Z + 2, N - 2) + BE(Z, N - 1)
$$

- BE(Z + 1, N - 2) + BE(Z + 1, N)
- BE(Z + 2, N - 1). (7)

FIG. 8. The distribution of the number of atomic nuclei as functions of deviations between theoretical calculations and experimental data for binding energies. The nuclei for which the BE deviations fall below 0.7 MeV are divided into seven equally spaced groups. The normal distribution curves fit the number of nuclei for which the $\sigma_{\rm rms}$ is lower than 0.7 MeV. We note that 63(51.6%) of the nuclei in ANN2, 32(26.2%) of the nuclei in ANN4, 16(13.1%) of the nuclei in ANN6, and 7(5.7%) of the nuclei in ANN7 have deviations over 0.7 MeV.

By training 450 nuclei randomly chosen in AME2016 with $Z \geqslant 20$, the root-mean-square deviation of the predictions on the training set was 0.186 MeV and 0.316 MeV for heavier nuclei with $A \ge 50$.

It is interesting to check how adding the GK relation affects the performance of our ANNs. For such a purpose, we designed two training methods. The first training method is to train 20% of the AME2016 (686 nuclei, randomly chosen) and take the remaining 80% (2748 nuclei) and the 122 nuclei (test20) of AME2020 as the test sets. This method is similar to MDN since it was also trained using about 20% nuclei of AME2016.

Note that our training set is not the same as the training set of Ref. $[55]$. From Table IV, it can be seen that the performance of ANN7 with the constraint of the GK relation is better than the performance of ANN7 without such a constraint not only on the training set but also on the two test sets.

TABLE IV. Performance of ANN7 with or without the constraint of the GK relation by using only 20% of the nuclear masses compiled in AME2016 for training, in comparison with the MDN method [\[55\]](#page-8-0). The rms deviations are given in units of MeV.

Method	training set	test16	test20
ANN7 without GK	0.65	1.02	1.54
ANN7 with GK	0.23	0.42	0.94
MDN	0.67	0.76	1.08

TABLE V. Performance of ANNs with and without the constraint of the GK relation.

Model	Without the GK constraint		With the GK constraint	
	training set	test set	training set	test set
ANN7	0.18	0.34	0.23	0.26

On the other hand, MDN performs worse than ANN7 on both test16 and test20. 3

The second method is to apply the constraint of the GK relation on our ANN7. Table V shows the results without and with the constraint of the GK relation. It is clear that adding the GK constraint can result in a slight improvement in the performance on the test set but no sign of improvement on the training set. This is in sharp contrast with the first test we performed using only 20% of the total nuclei contained in AME2016 as the training set.

From the above studies we performed, we conclude that using a subset of the masses contained in AME2016 as the training set, adding the GK relation can be beneficial, as demonstrated in Ref. [\[55\]](#page-8-0). On the other hand, using all the masses contained in AME2016 as the test set, as we did in this work, adding the GK relation has a minor impact on the performance of the ANNs. This corroborates the intuitive understanding that adding constraints goes in the same direction as adding more features and can be beneficial for scenarios where data are limited.

IV. SUMMARY AND OUTLOOK

In this work, we developed a deep neural network with seven physically motivated features: Z , N , Z_{EO} , N_{EO} , ΔZ , ΔN , and ASY. We studied the nuclear masses compiled in AME2020 (measured and extrapolated), achieving a description with a root-mean-square deviation around 0.2 MeV which is much smaller than the previous work [\[37\]](#page-8-0) and closer to those of Refs. [\[55,56\]](#page-8-0). The success of our work further demonstrated the importance of considering relevant physical information, i.e., "feature engineering," when applying machine learning methods to study systems for which only limited data are available.

³We used the numerical results provided to us in $[61]$.

It is interesting to note that the description of the nuclear binding energies achieved in the present work is similar to those of Refs. [\[55,56\]](#page-8-0), but our work differs from those of Refs. [\[55,56\]](#page-8-0) in many details: the networks, constraints, and input features. In Ref. [\[56\]](#page-8-0), 12 features are used. In our approach, we found that only six of them $(Z, N, Z_{EO}, N_{EO}, \Delta Z,$ ΔN) are relevant. On the other hand, Ref. [\[55\]](#page-8-0) considered the constraint of the GK relation in addition to eight features. Nevertheless, the similar results in these works support the conclusion that machine learning methods are powerful enough to predict nuclear binding energies at a level comparable to or even better than the most refined conventional theoretical models.

This work reveals that for systems with limited data, considering input features containing the most relevant physical information can be critical to the success of physical studies using machine learning methods. Turning the argument around, by trial and error, one can also anticipate the discovery of "new physics" by examining the deficiency of ANNs in describing such systems.

We note that the neural network developed in the present work can already describe experimental binding energies with a precision of 300 keV, reaching a level comparable to the global mass formula WS4 [\[22\]](#page-8-0) or even better than the much more sophisticated deformed relativistic Hartree-Bogoliubov theory in continuum [\[31\]](#page-8-0). As a result, the results can provide alternative theoretical binding energy inputs to various nuclear physics and nuclear astrophysics studies where experimental data are unavailable. For instance, one can use the predictions to study *r*-process nucleosynthesis [\[62\]](#page-8-0) or predict the halflives of potential proton emitters [\[63\]](#page-8-0), which is helpful for planning future experiments.

ACKNOWLEDGMENTS

We want to thank E. Yüksel and M. R. Mumpower for the useful communications and for sending us their data. This work is partly supported by the National Natural Science Foundation of China under Grants No. 11735003, No. 11975041, and No. 11961141004. X.-X.D. is supported by the Academic Excellence Foundation of BUAA for PhD Students and the China Scholarship Council.

- [1] V. Zelevinsky, [Acta Phys. Pol. A](https://doi.org/10.12693/APhysPolA.128.1008) **128**, 1008 (2015).
- [2] A. de Roubin *et al.*, Phys. Rev. C **96**[, 014310 \(2017\);](https://doi.org/10.1103/PhysRevC.96.014310) **97**, [059902\(E\) \(2018\).](https://doi.org/10.1103/PhysRevC.97.059902)
- [3] Q. Mo, M. Liu, and N. Wang, Phys. Rev. C **90**[, 024320 \(2014\).](https://doi.org/10.1103/PhysRevC.90.024320)
- [4] M. Rosenbusch *et al.*, Phys. Rev. Lett. **114**[, 202501 \(2015\).](https://doi.org/10.1103/PhysRevLett.114.202501)
- [5] [D. Lunney, J. M. Pearson, and C. Thibault,](https://doi.org/10.1103/RevModPhys.75.1021) Rev. Mod. Phys. **75**, 1021 (2003).
- [6] W. S. Porter *et al.*, Phys. Rev. C **106**[, 024312 \(2022\).](https://doi.org/10.1103/PhysRevC.106.024312)
- [7] T. Tanaka *et al.*, Phys. Rev. Lett. **124**[, 052502 \(2020\).](https://doi.org/10.1103/PhysRevLett.124.052502)
- [8] M. E. Burbidge, G. R. Burbidge, W. A. Fowler, and F. Hoyle, [Rev. Mod. Phys.](https://doi.org/10.1103/RevModPhys.29.547) **29**, 547 (1957).
- [9] A. Boehnlein *et al.*, [Rev. Mod. Phys.](https://doi.org/10.1103/RevModPhys.94.031003) **94**, 031003 (2022).
- [10] M. R. Mumpower, R. Surman, G. C. McLaughlin, and A. Aprahamian, [Prog.](https://doi.org/10.1016/j.ppnp.2015.12.002)[Part.](https://doi.org/10.1016/j.ppnp.2015.12.002)[Nucl.](https://doi.org/10.1016/j.ppnp.2015.12.002)[Phys.](https://doi.org/10.1016/j.ppnp.2015.12.002) **86**, 86 (2016); **87**, 116(E) (2016).
- [11] H. Schatz and W.-J. Ong, [Astrophys. J.](https://doi.org/10.3847/1538-4357/aa7de9) **844**, 139 (2017).
- [12] [P. Möller, A. J. Sierk, T. Ichikawa, and H. Sagawa,](https://doi.org/10.1016/j.adt.2015.10.002) At. Data Nucl. Data Tables **109–110**, 1 (2016).
- [13] X. W. Xia *et al.*, [At. Data Nucl. Data Tables](https://doi.org/10.1016/j.adt.2017.09.001) **121–122**, 1 (2018).
- [14] [W. Zhang, Z. Li, W. Gao, and T. T. Sun,](https://doi.org/10.1088/1674-1137/ac7b18) Chin. Phys. C **46**, 104105 (2022).
- [15] M. Wang, W. J. Huang, F. G. Kondev, G. Audi, and S. Naimi, Chin. Phys. C **45**[, 030003 \(2021\).](https://doi.org/10.1088/1674-1137/abddaf)
- [16] W. Nazarewicz, Nat. Phys. **14**[, 537 \(2018\).](https://doi.org/10.1038/s41567-018-0163-3)
- [17] [L.-S. Geng, H. Toki, and J. Meng,](https://doi.org/10.1143/PTP.113.785) Prog. Theor. Phys. **113**, 785 (2005).
- [18] C. F. v. Weizsäcker, Z. Phys. **96**[, 431 \(1935\).](https://doi.org/10.1007/BF01337700)
- [19] N. Wang, M. Liu, and X. Wu, Phys. Rev. C **81**[, 044322 \(2010\).](https://doi.org/10.1103/PhysRevC.81.044322)
- [20] [N. Wang, Z. Liang, M. Liu, and X. Wu,](https://doi.org/10.1103/PhysRevC.82.044304) Phys. Rev. C **82**, 044304 (2010).
- [21] [M. Liu, N. Wang, Y. Deng, and X. Wu,](https://doi.org/10.1103/PhysRevC.84.014333) Phys. Rev. C **84**, 014333 (2011).
- [22] [N. Wang, M. Liu, X. Wu, and J. Meng,](https://doi.org/10.1016/j.physletb.2014.05.049) Phys. Lett. B **734**, 215 (2014).
- [23] [Y. Gambhir, P. Ring, and A. Thimet,](https://doi.org/10.1016/0003-4916(90)90330-Q) Ann. Phys. **198**, 132 (1990).
- [24] J. Duflo and A. P. Zuker, Phys. Rev. C **52**[, R23 \(1995\).](https://doi.org/10.1103/PhysRevC.52.R23)
- [25] [S. Goriely, N. Chamel, and J. M. Pearson,](https://doi.org/10.1103/PhysRevC.88.061302) Phys. Rev. C **88**, 061302(R) (2013).
- [26] [S. Goriely, N. Chamel, and J. M. Pearson,](https://doi.org/10.1103/PhysRevC.88.024308) Phys. Rev. C **88**, 024308 (2013).
- [27] [S. Goriely, S. Hilaire, M. Girod, and S. Peru,](https://doi.org/10.1103/PhysRevLett.102.242501) Phys. Rev. Lett. **102**, 242501 (2009).
- [28] [S. Goriely, N. Chamel, and J. M. Pearson,](https://doi.org/10.1103/PhysRevC.93.034337) Phys. Rev. C **93**, 034337 (2016).
- [29] [P. Moller, J. R. Nix, W. D. Myers, and W. J. Swiatecki,](https://doi.org/10.1006/adnd.1995.1002) At. Data Nucl. Data Tables **59**, 185 (1995).
- [30] [P. Möller, W. D. Myers, H. Sagawa, and S. Yoshida,](https://doi.org/10.1103/PhysRevLett.108.052501) *Phys. Rev.* Lett. **108**, 052501 (2012).
- [31] K. Zhang *et al.* [\(DRHBc Mass Table Collaboration\),](https://doi.org/10.1016/j.adt.2022.101488) At. Data Nucl. Data Tables **144**, 101488 (2022).
- [32] P. Harris *et al.*, [arXiv:2203.16255.](https://arxiv.org/abs/2203.16255)
- [33] [T. Bayram, S. Akkoyun, and S. O. Kara,](https://doi.org/10.1016/j.anucene.2013.07.039) Ann. Nucl. Energy **63**, 172 (2014).
- [34] [R. Utama, J. Piekarewicz, and H. B. Prosper,](https://doi.org/10.1103/PhysRevC.93.014311) Phys. Rev. C **93**, 014311 (2016).
- [35] R. Utama and J. Piekarewicz, Phys. Rev. C **96**[, 044308 \(2017\).](https://doi.org/10.1103/PhysRevC.96.044308)
- [36] R. Utama and J. Piekarewicz, Phys. Rev. C **97**[, 014306 \(2018\).](https://doi.org/10.1103/PhysRevC.97.014306)
- [37] [E. Yüksel, D. Soydaner, and H. Bahtiyar,](https://doi.org/10.1142/S0218301321500178) Int. J. Mod. Phys. E **30**, 2150017 (2021).
- [38] Z. M. Niu and H. Z. Liang, [Phys. Lett. B](https://doi.org/10.1016/j.physletb.2018.01.002) **778**, 48 (2018).
- [39] [X.-X. Dong, R. An, J.-X. Lu, and L.-S. Geng,](https://doi.org/10.1103/PhysRevC.105.014308) Phys. Rev. C **105**, 014308 (2022).
- [40] [X.-X. Dong, R. An, J.-X. Lu, and L.-S. Geng,](https://doi.org/10.1016/j.physletb.2023.137726) Phys. Lett. B **838**, 137726 (2023).
- [41] [D. Wu, C. L. Bai, H. Sagawa, and H. Q. Zhang,](https://doi.org/10.1103/PhysRevC.102.054323) Phys. Rev. C **102**, 054323 (2020).
- [42] [R. Utama, W.-C. Chen, and J. Piekarewicz,](https://doi.org/10.1088/0954-3899/43/11/114002) J. Phys. G **43**, 114002 (2016).
- [43] [C.-Q. Li, C.-N. Tong, H.-J. Du, and L.-G. Pang,](https://doi.org/10.1103/PhysRevC.105.064306) Phys. Rev. C **105**, 064306 (2022).
- [44] Z. M. Niu, H. Z. Liang, B. H. Sun, W. H. Long, and Y. F. Niu, Phys. Rev. C **99**[, 064307 \(2019\).](https://doi.org/10.1103/PhysRevC.99.064307)
- [45] [C.-W. Ma, D. Peng, H.-L. Wei, Y.-T. Wang, and J. Pu,](https://doi.org/10.1088/1674-1137/abb657) Chin. Phys. C **44**, 124107 (2020).
- [46] C.-W. Ma, D. Peng, H.-L. Wei, Z.-M. Niu, Y.-T. Wang, and R. Wada, Chin. Phys. C **44**[, 014104 \(2020\).](https://doi.org/10.1088/1674-1137/44/1/014104)
- [47] [Z.-A. Wang, J. Pei, Y. Liu, and Y. Qiang,](https://doi.org/10.1103/PhysRevLett.123.122501) Phys. Rev. Lett. **123**, 122501 (2019).
- [48] M. Carnini and A. Pastore, J. Phys. G **47**[, 082001 \(2020\).](https://doi.org/10.1088/1361-6471/ab92e3)
- [49] [Z. Gao, Y. Wang, H. Lü, Q. Li, C. Shen, and L. Liu,](https://doi.org/10.1007/s41365-021-00958-z) Nucl. Sci. Tech. **32**, 118 (2021).
- [50] H. F. Zhang, L. H. Wang, J. P. Yin, P. H. Chen, and H. F. Zhang, J. Phys. G **44**[, 045110 \(2017\).](https://doi.org/10.1088/1361-6471/aa5d78)
- [51] [X. H. Wu, Y. Y. Lu, and P. W. Zhao,](https://doi.org/10.1016/j.physletb.2022.137394) Phys. Lett. B **834**, 137394 (2022).
- [52] Z. M. Niu and H. Z. Liang, Phys. Rev. C **106**[, L021303 \(2022\).](https://doi.org/10.1103/PhysRevC.106.L021303)
- [53] [H. Bahtiyar, D. Soydaner, and E. Yüksel,](https://doi.org/10.1016/j.asoc.2022.109470) Appl. Soft Comput. **128**, 109470 (2022).
- [54] M. Wang, G. Audi, F. Kondev, W. Huang, S. Naimi, and X. Xu, Chin. Phys. C **41**[, 030003 \(2017\).](https://doi.org/10.1088/1674-1137/41/3/030003)
- [55] M. R. Mumpower, T. M. Sprouse, A. E. Lovell, and A. T. Mohan, Phys. Rev. C **106**[, L021301 \(2022\).](https://doi.org/10.1103/PhysRevC.106.L021301)
- [56] A. E. Lovell, A. T. Mohan, T. M. Sprouse, and M. R. Mumpower, Phys. Rev. C **106**[, 014305 \(2022\).](https://doi.org/10.1103/PhysRevC.106.014305)
- [57] D. Hendrycks and K. Gimpel, $arXiv:1606.08415$.
- [58] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (Curran Associates, Inc., Red Hook, NY, 2019), pp. 8024–8035.
- [59] D. P. Kingma and J. Ba, $arXiv:1412.6980$.
- [60] G. T. Garvey, W. J. Gerace, R. L. Jaffe, I. Talmi, and I. Kelson, [Rev. Mod. Phys.](https://doi.org/10.1103/RevModPhys.41.S1) **41**, S1 (1969).
- [61] M. R. Mumpower (private communication).
- [62] B. Sun, F. Montes, L. S. Geng, H. Geissel, Y. A. Litvinov, and J. Meng, Phys. Rev. C **78**[, 025806 \(2008\).](https://doi.org/10.1103/PhysRevC.78.025806)
- [63] Y. Xiao, S.-Z. Xu, R.-Y. Zheng, X.-X. Sun, L.-S. Geng, and S.-S. Zhang, Phys. Lett. B **845**[, 138160 \(2023\).](https://doi.org/10.1016/j.physletb.2023.138160)