

Deep learning approach to nuclear masses and α -decay half-lives

Chen-Qi Li*

Key Laboratory of Quark & Lepton Physics (MOE) and Institute of Particle Physics, Central China Normal University, Wuhan 430079, China
and Physics Department, University of California, Berkeley, California 94720, USA

Chao-Nan Tong, Hong-Jing Du, and Long-Gang Pang[†]

Key Laboratory of Quark & Lepton Physics (MOE) and Institute of Particle Physics, Central China Normal University, Wuhan 430079, China



(Received 2 March 2022; revised 8 May 2022; accepted 31 May 2022; published 17 June 2022)

Ab initio calculations of nuclear masses, binding energy, and α -decay half-lives are intractable for heavy nucleus because of the curse of dimensionality in many-body quantum simulations as proton number (N) and neutron number (Z) grow. We take advantage of the powerful nonlinear transformation and feature representation ability of deep neural network (DNN) to predict the nuclear masses and α -decay half-lives. For nuclear binding energy prediction problem we achieve standard deviation $\sigma = 0.263$ MeV on 10-fold cross validation on 2149 nuclei. Word-vectors which are high-dimensional representation of nuclei from the hidden layers of mass-regression DNN help us to calculate α -decay half-lives. For this task, we get $\sigma = 0.797$ on 100 times 10-fold cross validation on 350 nuclei on $\log_{10}T_{1/2}$ and $\sigma = 0.731$ on 486 nuclei. DNN is also used to reduce the residual of three-parameter Gamow formula on 159 even-even nuclei, from 0.3627 to 0.2297 on $\log_{10}T_{1/2}$, using 100 times 10-fold cross validation. We find physical *a priori* such as shell structure, magic numbers and augmented inputs inspired by finite-range droplet model are important for this small data regression task.

DOI: [10.1103/PhysRevC.105.064306](https://doi.org/10.1103/PhysRevC.105.064306)

I. INTRODUCTION

Ground-state nuclear mass (binding energy), α -decay half-life, and β -decay half-life are all important properties of the nucleus [1,2]. Accurately calculating and predicting these quantities are crucial for justifying *ab initio* quantum many-body calculations as well as phenomenology models, such as liquid droplet model and shell model. Understanding the creation of heavy elements in our universe by the rapid neutron-capture process or r -process require accurate nuclear mass predictions in nuclear astrophysics [3,4]. These quantities also play important roles in nuclear stability studies which may guide us to find more super-heavy nuclei and even the “super-heavy island.”

Machine learning is a collection of algorithms that let the computer learn patterns from big data by themselves. The learned patterns are widely used in classification and regression tasks to get the state-of-the-art performance in many scientific problems. Various machine learning methods have been used in nuclear physics for regression tasks. For example, Bayesian neural network [5–8] (BNN), radial basis function (RBF), light gradient boosting machine [9] (LightGBM), and many other methods [10,11] have been applied to predict the residual between true nuclear mass and phenomenology models, such as FRDM [12], Weizsäcker-Skyrme (WS) mass model [13,14], and Skyrme

HartreeFock-Bogoliubov (HFB) model [15,16]. These machine learning methods have improved both accuracy and extrapolation ability greatly [17,18]. The mass predicted using Machine Learning algorithm is used to construct the outer crust equation of state (EoS) of a neutron star which is comparable to existing models [19].

Deep neural network (DNN) with multiple hidden layers is the most popular machine learning tool, which has outperformed many traditional ways and advances many scientific researches to the state-of-the-art. Read Refs. [20,21] for more detailed reviews on its applications in nuclear physics. Deep neural network is proved to have universal approximation ability with at least one hidden layer. It has long been used to predict the ground-state mass of nuclei [22–25]. However, even with the rapid development in recent years, the performance of deep neural network on ground state nuclear mass are still much worse than other methods, as shown in many recent researches. For example, the root-mean-square (rms) deviation is above 1 MeV using a four-layer neural network [26] recently. As a comparison, the rms deviation from improved WS model is 0.336 MeV [14]. The best performance from a decision tree-based method LightGBM has rms deviation around 0.170 MeV [9]. We believe that the feed-forward neural network used in recent studies are not deep and wide enough to reach its best performance. After performing a simple neural architecture search (NAS), we achieve rms deviation 0.263 MeV in 10-fold cross validation.

The performance of deep neural network highly depends on the amount of data, however, there are only about 2500 existing experiment data till 2020 [27] for nuclear mass

*steveli@mails.ccnu.edu.cn

†lgpang@ccnu.edu.cn

regression, and less than 500 available experiment data for α -decay half-life regression. One important question is whether a deep neural network can learn from this tiny data without terrible over-fitting. Will the patterns learned in mass prediction help the α -decay half-life prediction?

One method is to train the same network to achieve multiple tasks, which is called multi-task learning (MTL). In this way, different tasks have shared modules in the network as well as their own modules. MTL optimizes the shared parameter with limited data for each task. MTL has been used to describe the giant dipole resonance key parameters [28].

Another method is called representation learning. We use the latent representations of each nucleus learned in the nuclear mass prediction task to assist the α -decay half-life prediction. The latent representation is similar to word-vector in natural language processing tasks where each word is represented by a high-dimensional vector with 256 or 512 floating numbers, to represent individual words in a text, taking into account the context and other surrounding words learned by the deep neural network in other big-data problems. Using a 256-dimensional word vector from the previous nuclear mass prediction task as a new representation of a nucleus, we predict the α -decay half-life and verify that the word-vector really improve its performance.

The paper is organized as follows: In Sec. II we introduce the network structure, the input data structure and the prediction accuracy of nuclear binding energy. In Secs. III and IV we introduce the method of nuclear-representation and its application in predicting the α -decay half-lives. The discussion and summary will be given in Secs. V and VI.

II. NUCLEAR BINDING ENERGY PREDICTION

A. Methods

The nuclear binding energy prediction is a supervised regression problem. The objective is to minimize the the residual which is defined as the difference between experimental data and semiempirical models. Two semiempirical models are used, one is Bethe-Weizsäcker model (BWM) and the other is liquid-droplet model (LDM). Both models have root-mean-square error larger than 2 MeV.

Two types of inputs are used. The first type consists of three native features, Z , N , and A , for each nucleus. The second type has 26 features (see the Appendix) with physical *a priori* as shown below. The neural network has one adjustable architecture whose number of layers are $n + 2$ and neurons per hidden layer equal to m or $4 \times m$.

Data flow in the feed-forward neural network according to the following equation for adjacent layers,

$$h_i = \sigma \left(\sum_j w_{ij} x_j + b_i \right), \quad (1)$$

where h_i represents the value of the i th neuron in the next layer, and x_j represents values of the j th neuron in the previous layer. The network parameters are w_{ij} (weights) and b_i (bias), which are initialized with random numbers and are adjusted gradually during training using stochastic gradient descent algorithm. The feature vector in the previous layer

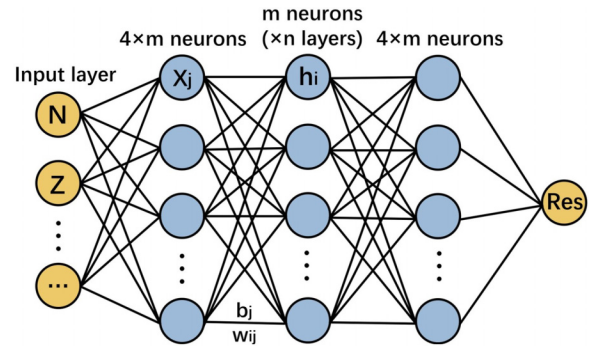


FIG. 1. The adjustable neural network structure with the number of hidden layers changed by n and the width of hidden layers changed by m . The output layer is the residual defined in Eq. (2).

are first linearly transformed through $z_i = x_j w_{ij} + b_i$, and then feed to a nonlinear activation function $h_i = \sigma(z_i)$. The operation corresponds to a manipulations of feature vector in high-dimensional space.

As shown in Fig. 1, the input consists of three native features (Z , N , A) or 26 physics-informed features. In the output layer, there is one neuron representing the residual binding energy. In between, there are $n + 2$ hidden layers. The first and the last hidden layer have $4 \times m$ neurons and the other n hidden layers have m neurons per layer. The number m is defined as the “width” of the neural network. The performance of the network prediction is scanned using 10-fold cross validation, for $n = (0, 2, 4, 6, 8, 10, 16)$ and $m = (8, 64, 256, 512)$.

Other special method used in our DNN: between each layer, batch-normalization method [29] is adopted to accelerate learning as well as avoid vanishing gradient and exploding gradient.

B. Performance scan and prediction accuracy

Shown in Fig. 2 are the 10-fold average RMS error for different numbers of hidden layers and different numbers of neurons per hidden layer. Using 26 features in the input, the RMS error decreases from 1.5 to 0.3 MeV as the width of the network increases from $m = 8$ to $m = 64$. Increasing m from 256 to 512 does not bring further improvement. However, the performance is not sensitive to the depth of the neural network. The RMS error changes slightly as one increases the number of hidden layers from 2 to 18. For $m = 256$, the RMS error reaches its minimum for 10 hidden layers ($n = 8$).

The parameters of the neural network are initialized with random numbers and adjusted during the training process, as a result, the final performance relies on the starting point of optimization in the parameter space. At rare times, the parameter optimization starts at a position which produces worse results than general cases. This happens in the performance scan for a network with structure ($n = 0, m = 256$). Although it is a rare event, we do not retrain the network for cherry picking. It reminds us that repeating the training process many times is a good way to capture the uncertainty of the network on small data.

According to the performance scan, the optimal network structure is ($n = 8, m = 256$) for LDM mass residual. As

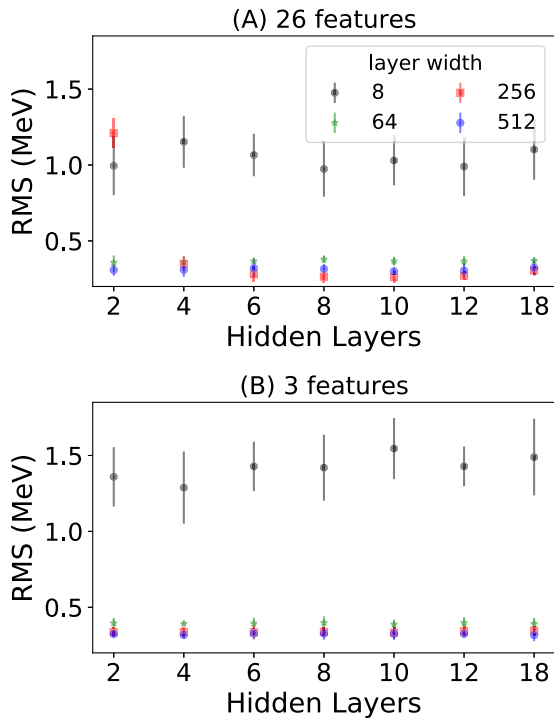


FIG. 2. The performance of the atomic mass prediction using 10-fold cross validation. The bands represent the range of root-mean-square error using different numbers of hidden layers and different numbers of neurons per layer in the architecture of the deep neural network. The optimal RMSE is around 0.22 MeV using 10 hidden layers with 1,024 neurons in the first and the last hidden layer and 256 neurons for eight other hidden layers, which correspond to width = 256 in the plot.

shown in Fig. 3, the average RMS error of 10-fold is approximately 263 keV for 26 features and 332 keV for 3 features. Using physical *a priori* as input, the RMS error reduces by 69 keV.

Shown in Fig. 4 are comparisons between semiempirical models and network predictions with 3 and 26 features. As the network is trained to predict the residual of semiempirical model, the new prediction error is defined as

$$\text{Residual} = M_{\text{exp}} - M_{\text{LDM}} - R_{\text{Network}}. \quad (2)$$

In this comparison, both training and testing data are included as what has been done in semiempirical models.

Shown in Fig. 5 is the comparison between network prediction and LDM for 322 new elements that have never been used to train the network or to fit the parameters of LDM. The network prediction for these 322 new elements has RMS error 0.605 MeV which is larger than the validation accuracy during training. However, it is much smaller than LDM where the RMS is around 2.542 MeV. Usually it is believed that a theoretical model with a few parameters generalize better than a deep neural network, because the later has millions of parameters (here our DNN has about one millions trainable parameters) and was thought to be easy to overfit to the training data and fail to extrapolate to new data. In this study, it is

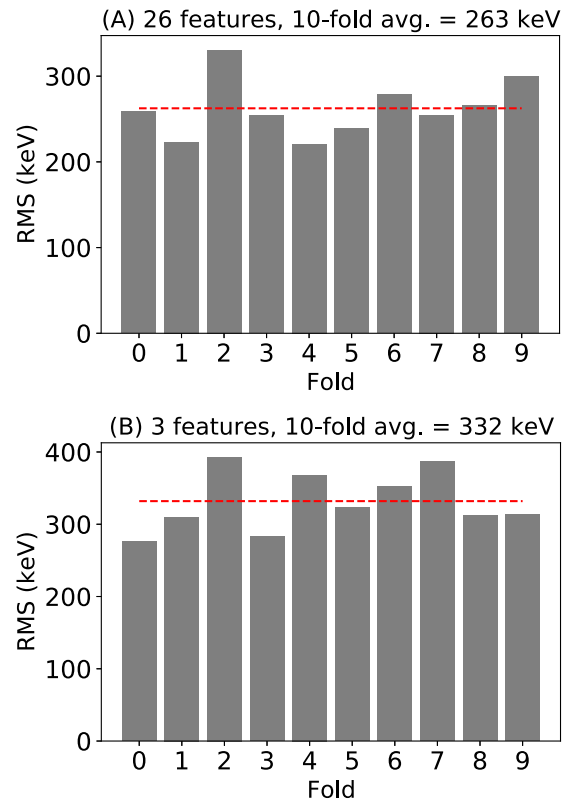


FIG. 3. The optimal performance of the atomic mass prediction using 10-fold cross validation. The RMS error of the 10-fold cross validation for (A) 26 features and (B) 3 features as the input of the optimal network ($n = 8$, $m = 256$).

shown that the deep learning generalizes better than LDM on new data.

III. GLOBAL α -DECAY HALF-LIFE PREDICTION

A. Methods

The neural network trained in the nuclear mass prediction task can help the α -decay half-lives [30,31] prediction in two folds. First, for super-heavy nucleus [32] whose Q value has no experimental measurements or *ab initio* calculations, the high precision network prediction provides a cheap way to compute the Q value [33], using the mass of the mother nucleus, daughter nucleus and the α particle. Second, the network trained in nuclear mass prediction produces a word-vector representation for each nucleus, it is simply get from one of its hidden layers with a high dimension. The representation encodes high-dimensional information of the nucleus which may help many other calculations in nuclear physics, such as α -decay and β -decay [34,35] half-lives or charge radius prediction. In the present work, we test the effect on α -decay half-lives prediction.

The Q value is a key information in α -decay and the half-lives is really sensitive to it according to most semiempirical formulas, e.g., Royer formula [36]. It will be an important feature to be used as the input of the neural network for α -decay half-lives prediction. Shown in Fig. 6 is the prediction

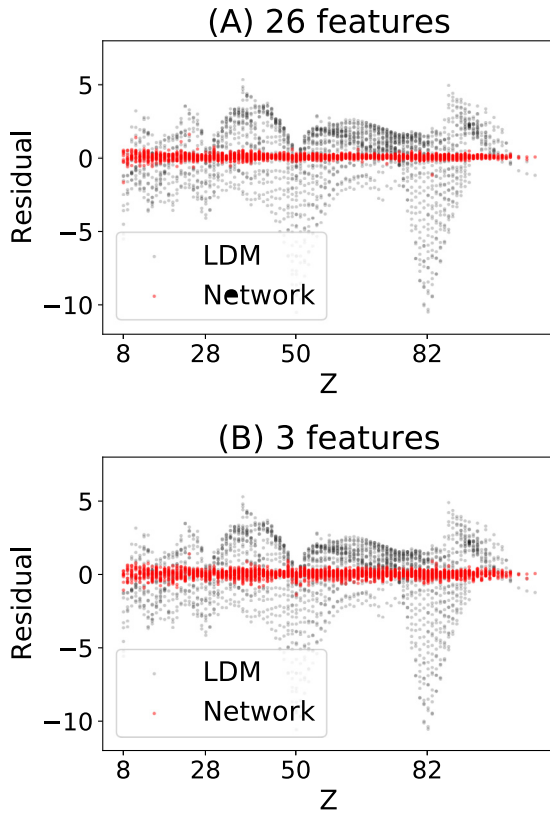


FIG. 4. The prediction error for all the nuclei as compared with LDM. Both training and testing data are used with (A) 26 features and (B) 3 features as the input of the optimal network ($n = 8$, $m = 256$).

error of the Q value as compared with experimental data for 486 nuclei. The performance is not good for some super-heavy nuclei, however, the average rms deviation is only 0.15 MeV. Figure 7 is a more detailed comparison for nuclei with proton number = 86, 87, 88, 89, 92, 93.

The network we used is smaller in the α -decay half-lives prediction after the performance scan. The structure is (n input, 128, 256, 256, 256, 256, 1) where “tanh” activation

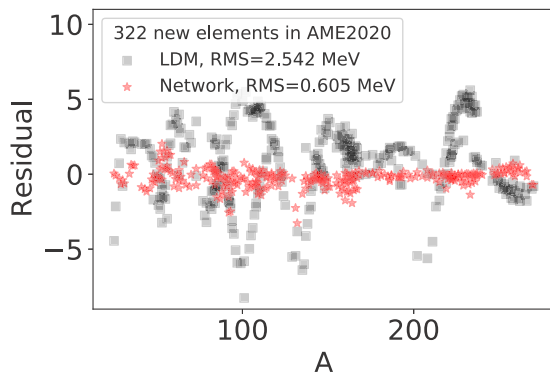


FIG. 5. The mass-residual prediction error for 322 new elements in AME2020 as compared with LDM.

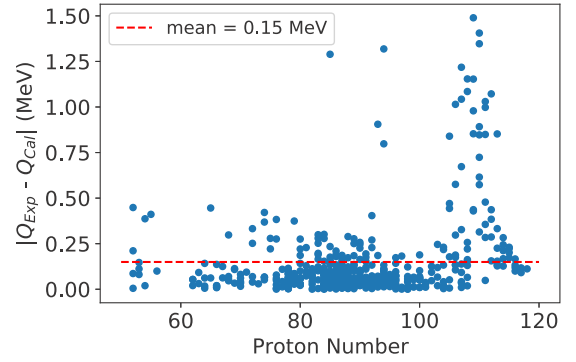


FIG. 6. The Q -value prediction error for 486 nuclei.

functions are used for the first and the last hidden layer, and “relu” activation functions are used for other hidden layers.

The training data used to get the nuclear word-vector is not those best physical nuclear mass models, which contain both macro and micro parts whose residual is only about 0.4 MeV [37–39]. We only use the macro part of Bethe-Weizsäcker model (BWM) and liquid-droplet model (LDM). In our experience, subtracting both macro and micro parts will not benefit

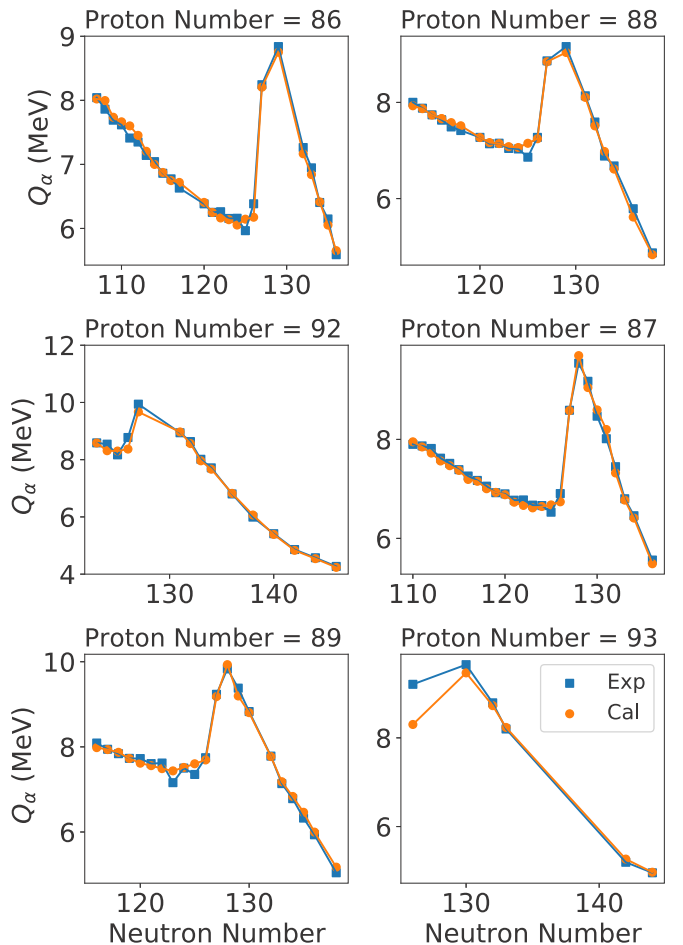


FIG. 7. The Q value from deep neural network (solid circles) as compared with experimental measurements (square) for nuclei with proton number = 86, 87, 88, 89, 92, 93.

nuclear word-vector learning. We observe that keeping the micro parts in the mass residual helps to reduce the α -decay half-lives predicting error. Although the learning is more difficult if the effects of micro parts are included in the mass residual, the network will try its best to encode the associated quantum properties of nucleus to minimize the differences to the residual in supervised learning. The encoded quantum properties in the nuclear word-vector helps the α -decay half-lives prediction.

The prediction accuracy for two types of inputs are calculated. The first type mainly consists of native features. The second type uses representations of nucleus learned in the atomic mass prediction task, which is the word-vector.

In the 10-fold cross validation, the data-set is evenly divided into 10 folds, 9 folds are used for training and 1 for validation. Since our training data set is small, even 10-fold cross validation method still have big fluctuation and cannot evaluate the performance in a credible way, so we do 100 times 10-fold cross validation to better evaluate the performance our neural network. As a result, there are 1000 validation scores for each type of inputs.

B. Results

The latent representations of nuclei are learned in the nuclear mass prediction task. Analogous to the spatial representation and momentum representation, the word-vector representations carry the ground-state information of nuclei using a high-dimensional array of floating numbers. The high-dimensional word-vector of nuclei are believed to capture physics of the many-body quantum system which will help other relating tasks. This representation helps the network to predict the α -decay half-lives better as compared with that trained with native features as inputs on the 350 nuclei, as shown in Fig. 8. We also test the performance of word-vector gotten from different hidden layers, as show in Fig. 9 and find that as the network goes deeper, the performance turns worse. The word vectors from deep layers seem to encode more information about the nuclear mass, which is objective of the pretraining task and will not help too much in other tasks. Word-vectors get from shallow hidden layers seem to be good representations for new tasks.

It might seem a bit confusing in 10, some histograms have “two peaks,” that is because there are two nuclei in the data set hard to be predicted ($Z = 64, N = 84$ and $Z = 71, N = 82$), so the folds that have these two will have bad scores.

If a nucleus in the α -decay table has not been used to train the mass prediction network, then its latent representation will not be as good as those nuclei in the training dataset. In the previous data-set with 350 nuclei, 26% are missing from the pretraining data. In another data-set which has 486 nuclei along with the experimental Q value, 29% have not been used to train the mass-residual prediction network. For this new data-set, training with 64 native features still performs better than the best word-vector performance. Using the 64 native features as inputs, as show in Fig. 10 we get an RMS = 0.7315 on 100 times 10-fold cross validation, and the average training loss is about 0.1.

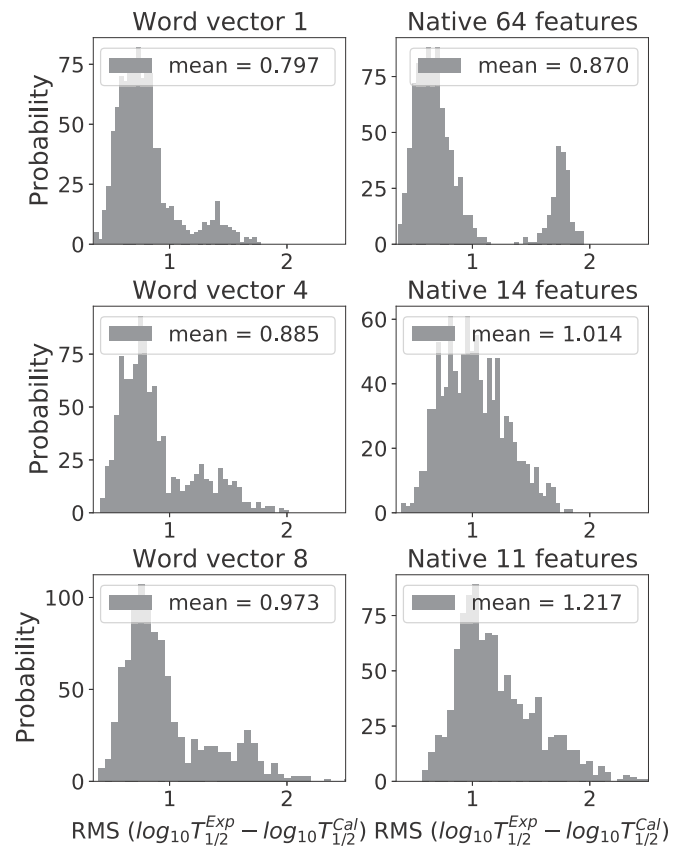


FIG. 8. Prediction for α -decay half-lives using native inputs as compared with word-vector representations on 350 nuclei. Word vector 1 means it is obtained from the first hidden layer of the deep neural network we used to predict nuclear mass; for the details of those native features, see the Appendix.

The division of training data and testing data can hugely influence the test result as we have seen in Fig. 10, the result of 100 times 10-fold cross validation. Figure 11 shows a certain division which lead to the best testing result.

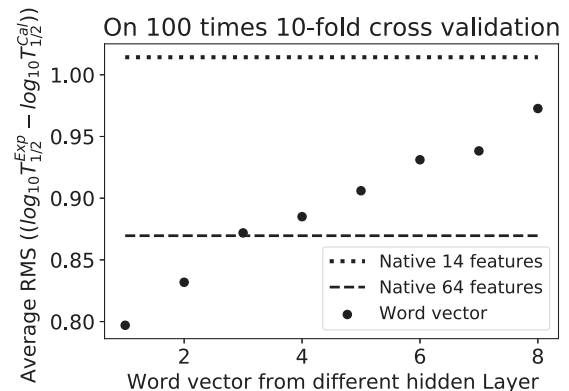


FIG. 9. The performance of word-vector from different hidden layers compared with native features as inputs.

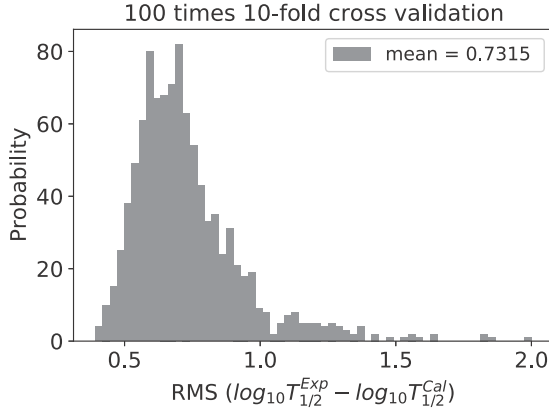


FIG. 10. Result of α -decay half-lives prediction on 486 nuclei using 64 native features as input.

IV. α -DECAY HALF-LIVES PREDICTION ON EVEN-EVEN NUCLEI

Three-parameter Gamow formula [40–42] has a quite low prediction error for the half-lives of even-even nuclei,

$$\log T = a \frac{Z}{Q} + b\sqrt{Z} + c. \quad (3)$$

The residual between Gamow formula and measurements can be further reduced either using a polynomial fit or a neural network. Figure 12 shows the performance of DNN in improving the residual. Using DNN, we can reduce the residual of Gamow formula from 0.3627 to 0.2297, on 100 times 10-fold cross validation. The inputs (native 64 inputs) and DNN structure is the same as what we use for the global half-life prediction.

Using polynomial fit, the lowest residual we can get is 0.3052 using the same cross validation. The DNN generalizes better than a polynomial fit even on this small data problem. We also try to fit the network prediction using polynomial functions to see if it can generate some analytical relation. And we find among 1 to 10 order polynomials, the second order can do the best job with a result of 0.2679 on 100 times 10-fold cross validation. Although it is still not as good as DNN's 0.2297, it is already much better than fit the Gamow

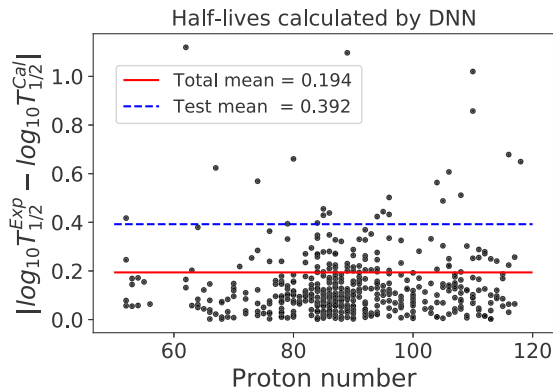


FIG. 11. Best network model prediction on 486 α -decay half-lives data.

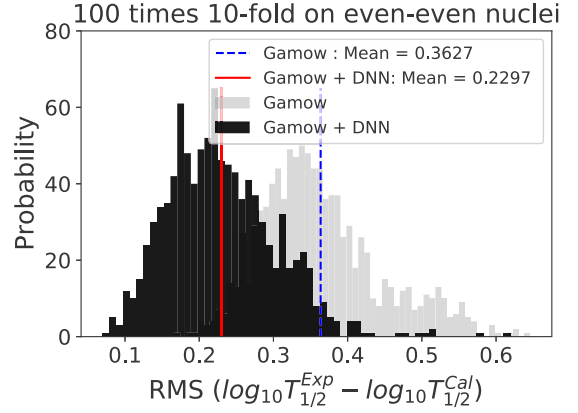


FIG. 12. The distribution and the mean residual of the Gamow formula as compared with the DNN improvement, on the α -decay half-lives of even-even nuclei.

residual directly, with lowest residual 0.3052 using different orders of polynomial functions. Higher order polynomials make the performance worse because of over fitting. The extracted coefficients are as follows:

$$\begin{aligned} &0.1547Z - 0.0222N - 0.4344Q \\ &- 0.0022Z^2 - 0.0008N^2 + 0.0204Q^2 \\ &+ 0.0024ZN + 0.0037ZQ - 0.0022NQ. \end{aligned} \quad (4)$$

As Z and N are much larger than Q , the following four terms, Z , Z^2 , ZN , and Q , have large contributions. Adding these four terms may improve the performance of the Gamow formula,

$$\log T = a \frac{Z}{Q} + b\sqrt{Z} + cZ^2 + dZN + eQ + fZ + g. \quad (5)$$

The cross validation performance of this modified Gamow formula on even-even nuclei is shown in Fig. 13.

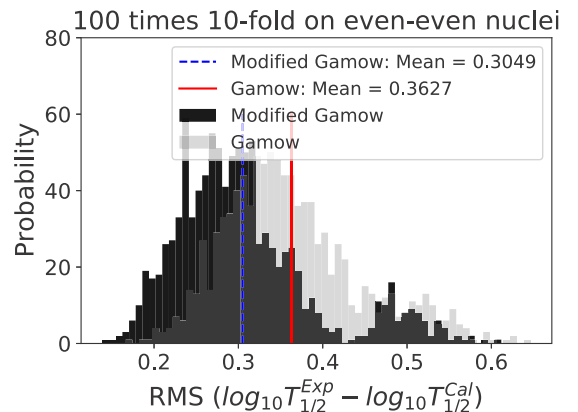


FIG. 13. The distribution and the mean residual of modified Gamow formula on the α -decay half-lives of even-even nuclei.

TABLE I. The correlations between various features and mass residual for LDM and BWM. According to the table, the magnitude of the correlation is strong for shell structure and magic numbers.

Features	Correlation	Features	Correlation
residual_LDM	1.000000	residual_BWM	1.000000
N_shell8	0.194654	N_shell8	0.380971
Z_shell7	0.120731	Z_shell7	0.363741
N_shell3	0.116282	Z_shell6	0.170536
Z_shell3	0.086444	N_shell7	0.143813
N_shell4	0.072662	Z	0.130579
Z_shell4	0.070161	A	0.104727
N_shell5	0.040211	N	0.086945
pair_energy	0.013236	$A^{2/3}$	0.082608
Z	0.002454	Z_shell3	0.039904
$A^{2/3}$	-0.000487	N-Z	0.005158
A	-0.007885	pair_energy	-0.003228
N	-0.014537	Z_shell4	-0.018830
Z_shell5	-0.028344	$A^{-1/3}$	-0.021580
Z_valence	-0.033500	N_shell3	-0.039164
Z_shell6	-0.036544	N_shell6	-0.039891
$A^{-1/3}$	-0.036847	Z_valence	-0.043764
N-Z	-0.041295	N_shell4	-0.054386
N_shell6	-0.044589	Z_shell5	-0.055684
N_shell7	-0.096653	N_shell5	-0.064732
magic_Z	-0.188633	N_valence	-0.153176
N_valence	-0.249767	magic_Z	-0.157655
magic_N	-0.253896	magic_N	-0.201970

V. DISCUSSION

Correlation matrix between the mass residual and features

The Pearson correlation coefficient is a good measure of the importance of each feature for the mass residual prediction, it also tells what physics is missing or not fully considered in the semiempirical model. The formula is given by

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

where r is the Pearson correlation coefficient, x and y represent values of two features for various samples, and n is the total number of samples. Shown in Table I are the correlations between mass residual and various features. Some common features are important for the LDM and BWM mass residual prediction, e.g., the number of neutrons on the 8th shell [43], the number of protons on the 7th shell, the number of valence neutrons, and whether N or Z are magic numbers. These features are important for both mass residual from LDM and BWM. In principle, the deep neural network is able to construct these features using native ones (N , Z , A). In practice, the data are small and the network may arrive at the final conclusion using other latent features. From the 10-fold cross validation, these features help to reduce the RMS error by 60 keV.

The correlation between mass residual and magic numbers is also shown in the heat-map Fig. 14. The deviations are large for either magic Z or magic N .

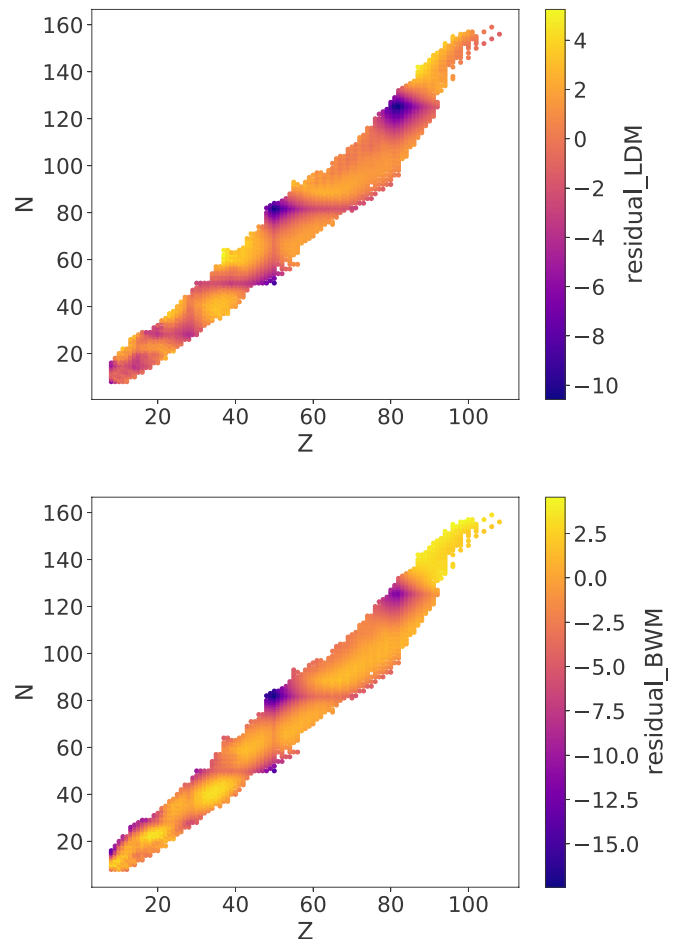


FIG. 14. The binding energy per nucleon as a function of proton number Z and neutron number N .

The correlation between binding energy per nucleon and number of valence nucleons are shown in Fig. 15. There seems to be several bands with each one shows a linear correlation between BE/A and valence nucleons. The band structure comes from different energy levels for different valence shells. In the differential study for different valence shells, light nuclei show negative correlation between BE/A and the number of valence nucleons as shown in Fig. 16, while heavy nuclei show positive correlation, as shown in Fig. 17.

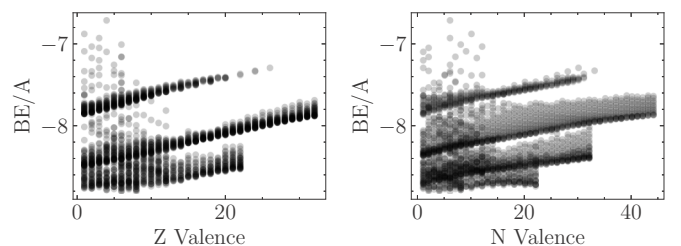


FIG. 15. The residual of binding energy per nucleon as a function of number of protons (left) and number of neutrons (right) on valence shells. Different shells contribute to different bands in the figure.

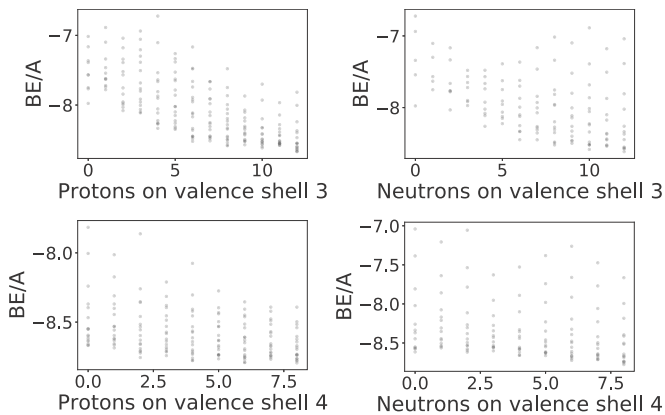


FIG. 16. The binding energy per nucleon as a function of number of protons (left) and number of neutrons (right) on valence shells 3 and 4.

VI. SUMMARY AND OUTLOOK

A deep neural network is trained to predict the nuclear mass residual between experimental data and phenomenological models. We achieve standard deviation $\sigma = 0.263$ MeV on 10-fold cross validation on 2149 nuclei. We verify that physical *a priori* (e.g., shell structure and magic numbers) helps to decrease the predicting error in this small data problem. The correlations between nuclear mass and various features, e.g., Z , N , magic number, as well as nucleons in each shell are calculated. It shows that the magic numbers as well as the number of nucleons on valence shells have strong correlation with mass residual. The values of neurons in the hidden layers of the network are used as latent representations or word-vectors of nuclei. These nuclear word-vectors from pretrained models in mass predicting task is used in a new α -decay half-lives prediction. We observe that keeping the micro part in the mass residual helps to learn a better word-vector of nuclei for α -decay prediction task. Word-vectors from shallow layers might better than deep layers indicating that deep layers might be more specific to the mass residual prediction.

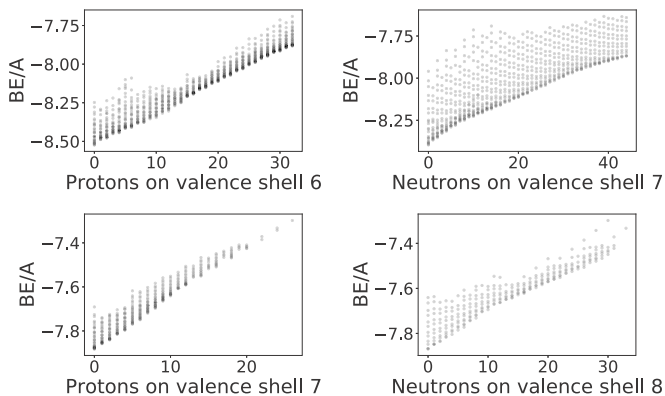


FIG. 17. The binding energy per nucleon as a function of number of protons (left) and number of neutrons (right) on valence shells of heavy nucleus.

In the future, the nuclear word-vector learned in the mass residual, α -decay half-lives prediction as well as other regression tasks can be used in relating tasks, such as β decay, r process, and so on. The method developed in the present paper paves a new way to use heterogeneous big data in the field of nuclear physics.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China under Grants No. 12075098 and No. 11861131009. Computations are performed at Nuclear Science Computer Center at CCNU (NSC3). L.G.P. also acknowledges the support provided by Huawei Technologies Co., Ltd.

The data used in this study are listed below:

- (1) 2149 nuclei (mass): FRDM (2012) [12]
- (2) 2471 nuclei (mass): AME2020 [27,44]
- (3) 350 nuclei (α decay): [45]
- (4) 486 nuclei (α decay): [46–48]

APPENDIX: FEATURES OF DIFFERENT INPUTS

Twenty-six features for nucleus:

- (1) (3 features) Z , N , A
- (2) (7 features) Number of protons on 7 shells
- (3) (8 features) Number of neutrons on 8 shells
- (4) (1 feature) Number of valence protons
- (5) (1 feature) Number of valence neutrons
- (6) (3 features) $N - Z$, $A^{2/3}$, $A^{-1/3}$
- (7) (1 features) Is Z a magic number? (1 for yes, 0 for no)
- (8) (1 features) Is N a magic number? (1 for yes, 0 for no)
- (9) (1 features) Pair energy: $\frac{(-1)^Z + (-1)^N}{2}$

Four features for odd-even:

- (1) (1 features) Is odd-odd nucleus? (1 for yes, 0 for no)
- (2) (1 features) Is even-even nucleus? (1 for yes, 0 for no)
- (3) (1 features) Is odd-even nucleus? (1 for yes, 0 for no)
- (4) (1 features) Is even-odd nucleus? (1 for yes, 0 for no)

Thirty features for nucleus:

- (1) (26 features mentioned above)
- (2) (4 features) odd-even

Native 11 features for α decay:

- (1) (9 features) Z , N , A for mother, daughter nucleus and He
- (2) (1 features) Q value calculated by network.
- (3) (1 features) $Q^{-1/2}$ calculated by network.

Native 14 features for α decay:

- (1) (9 features) Z , N , A for mother nucleus, daughter nucleus and He.
- (2) (1 features) Q value calculated by network.
- (3) (4 features) odd-even for mother nucleus.

Native 64 features for α decay:

- (1) (60 features) 30 features for mother nucleus and daughter nucleus.
- (2) (3 features) Z, N, A for α particle
- (3) (1 features) Q value calculated by network.

Word-vector 517 inputs for α decay:

- (1) (512 features) 256-dimensional word-vector features for mother nucleus and daughter nucleus.
- (2) (4 features) odd-even for mother nucleus.
- (3) (1 features) Q value calculated by network.

-
- [1] D. Lunney, J. M. Pearson, and C. Thibault, Recent trends in the determination of nuclear masses, *Rev. Mod. Phys.* **75**, 1021 (2003).
- [2] A. Akmal, V. R. Pandharipande, and D. G. Ravenhall, The equation of state of nucleon matter and neutron star structure, *Phys. Rev. C* **58**, 1804 (1998).
- [3] M. R. Mumpower, R. Surman, G. C. McLaughlin, and A. Arahamian, The impact of individual nuclear properties on r -process nucleosynthesis, *Prog. Part. Nucl. Phys.* **86**, 86 (2016).
- [4] D. Martin, A. Arcones, W. Nazarewicz, and E. Olsen, Impact of Nuclear Mass Uncertainties on the r -Process, *Phys. Rev. Lett.* **116**, 121101 (2016).
- [5] Z. M. Niu and H. Z. Liang, Nuclear mass predictions based on Bayesian neural network approach with pairing and shell effects, *Phys. Lett. B* **778**, 48 (2018).
- [6] U. B. Rodríguez, C. Z. Vargas, M. Gonçalves, S. B. Duarte, and F. Guzmán, Bayesian Neural Network improvements to nuclear mass formulae and predictions in the SuperHeavy Elements region, *Europhys. Lett.* **127**, 42001 (2019).
- [7] P. Möller, W. D. Myers, H. Sagawa, and S. Yoshida, New Finite-Range Droplet Mass Model and Equation-of-State Parameters, *Phys. Rev. Lett.* **108**, 052501 (2012).
- [8] X.-X. Dong, R. An, J.-X. Lu, and L.-S. Geng, Novel Bayesian neural network-based approach for nuclear charge radii, *Phys. Rev. C* **105**, 014308 (2022).
- [9] Z. Gao, Y. Wang, H. Lü, Q. Li, C. Shen, and L. Liu, Correction to: Machine learning the nuclear mass, *Nucl. Sci. Tech.* **32**, 118 (2021).
- [10] R. Utama and J. Piekarewicz, Refining mass formulas for astrophysical applications: A Bayesian neural network approach, *Phys. Rev. C* **96**, 044308 (2017).
- [11] Y. Liu, C. Su, J. Liu, P. Danielewicz, C. Xu, and Zhongzhou Ren, Improved naive Bayesian probability classifier in predictions of nuclear mass, *Phys. Rev. C* **104**, 014315 (2021).
- [12] P. Möller, A. J. Sierk, T. Ichikawa, and H. Sagawa, Nuclear ground-state masses and deformations: Frdm(2012), *At. Data Nucl. Data Tables* **109**, 1 (2016).
- [13] N. Wang, M. Liu, X. Wu, and J. Meng, Surface diffuseness correction in global mass formula, *Phys. Lett. B* **734**, 215 (2014).
- [14] M. Liu, N. Wang, Y. Deng, and X. Wu, Further improvements on a global nuclear mass model, *Phys. Rev. C* **84**, 014333 (2011).
- [15] S. Goriely, N. Chamel, and J. M. Pearson, Skyrme-Hartree-Fock-Bogoliubov Nuclear Mass Formulas: Crossing the 0.6 MeV Accuracy Threshold with Microscopically Deduced Pairing, *Phys. Rev. Lett.* **102**, 152503 (2009).
- [16] W. D. Myers, W. J. Swiatecki, T. Kodama, L. J. El-Jaick, and Eberhard R. Hilf, Droplet model of the giant dipole resonance, *Phys. Rev. C* **15**, 2032 (1977).
- [17] Z. M. Niu, J. Y. Fang, and Y. F. Niu, Comparative study of radial basis function and Bayesian neural network approaches in nuclear mass predictions, *Phys. Rev. C* **100**, 054311 (2019).
- [18] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, Bayesian approach to model-based extrapolation of nuclear observables, *Phys. Rev. C* **98**, 034318 (2018).
- [19] U. Murarka, K. Banerjee, T. Malik, and C. Providência, The neutron star outer crust equation of state: A machine learning approach, *J. Cosmol. Astropart. Phys.* **01** (2022) 045.
- [20] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [21] A. Boehnlein *et al.*, Artificial intelligence and machine learning in nuclear physics, [arXiv:2112.02309](https://arxiv.org/abs/2112.02309) (2021).
- [22] S. Gazula, J. W. Clark, and H. Bohr, Learning and prediction of nuclear stability by neural networks, *Nucl. Phys. A* **540**, 1 (1992).
- [23] S. Athanassopoulos, E. Mavrommatis, K. A. Gernoth, and J. Walter Clark, Nuclear mass systematics using neural networks, *Nucl. Phys. A* **743**, 222 (2004).
- [24] T. Bayram, S. Akkoyun, and S. Okan Kara, A study on ground-state energies of nuclei by using neural networks, *Ann. Nucl. Energy* **63**, 172 (2014).
- [25] R. Utama, J. Piekarewicz, and H. B. Prosper, Nuclear mass predictions for the crustal composition of neutron stars: A Bayesian neural network approach, *Phys. Rev. C* **93**, 014311 (2016).
- [26] E. Yüksel, D. Soydaner, and H. Bahtiyar, Nuclear binding energy predictions using neural networks: Application of the multilayer perceptron, *Int. J. Mod. Phys. E* **30**, 2150017 (2021).
- [27] M. Wang, W. J. Huang, F. G. Kondev, G. Audi, and S. Naimi, The AME 2020 atomic mass evaluation (II). Tables, graphs, and references, *Chinese Phys. C* **45**, 030003 (2021).
- [28] J. H. Bai, Z. M. Niu, B. Y. Sun, and Y. F. Niu, The description of giant dipole resonance key parameters with multitask neural networks, *Phys. Lett. B* **815**, 136147 (2021).
- [29] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in *Proceedings of the International Conference on Machine Learning (ICML)* (2015), pp. 448–456.
- [30] B. Alex Brown, Simple relation for alpha decay half-lives, *Phys. Rev. C* **46**, 811 (1992).
- [31] B. Buck, A. C. Merchant, and S. M. Perez, Half-lives of favored alpha decays from nuclear ground states, *At. Data Nucl. Data Tables* **54**, 53 (1993).
- [32] J. P. Cui, Y. L. Zhang, S. Zhang, and Y. Z. Wang, α -decay half-lives of superheavy nuclei, *Phys. Rev. C* **97**, 014316 (2018).
- [33] U. B. Rodríguez, C. Zuñiga Vargas, M. Gonçalves, S. Barbosa Duarte, and F. Guzmán, Alpha half-lives calculation of superheavy nuclei with Q_α -value predictions based on the Bayesian neural network approach, *J. Phys. G: Nucl. Part. Phys.* **46**, 115109 (2019).
- [34] N. J. Costiris, E. Mavrommatis, K. A. Gernoth, and J. W. Clark, Decoding β -decay systematics: A global statistical model for β^- half-lives, *Phys. Rev. C* **80**, 044332 (2009).

- [35] N. J. Costiris, E. Mavrommatis, K. A. Gernoth, J. W. Clark, and H. Li, Statistical global modeling of beta-decay half-lives systematics using multilayer feed-forward neural networks and support vector machines, [arXiv:0809.0383](https://arxiv.org/abs/0809.0383) (2008).
- [36] D. T. Akrawy, H. Hassanabadi, S. S. Hosseini, and K. P. Santhosh, Systematic study of α -decay half-lives using Royer and related formula, *Nucl. Phys. A* **971**, 130 (2018).
- [37] J. Toivanen, J. Dobaczewski, M. Kortelainen, and K. Mizuyama, Error analysis of nuclear mass fits, *Phys. Rev. C* **78**, 034306 (2008).
- [38] R. Utama and J. Piekarewicz, Validating neural-network refinements of nuclear mass models, *Phys. Rev. C* **97**, 014306 (2018).
- [39] K. P. Santhosh, Jayesh George Joseph, and B. Priyanka, Fine structure in the alpha-decay of odd-even nuclei, *Nucl. Phys. A* **877**, 1 (2012).
- [40] Gamow, *Zur Quantentheorie des Atomkernes* (1928).
- [41] R. H. Stuewer, *Gamow's Theory of Alpha-Decay* (Springer Netherlands, Dordrecht, 1986), pp. 147–186.
- [42] B. R. Holstein, Understanding alpha decay, *Am. J. Phys.* **64**, 1061 (1996).
- [43] D. N. Poenaru, R. A. Gherghescu, and N. Carjan, Alpha-decay lifetimes semiempirical relationship including shell effects, *Europhys. Lett.* **77**, 62001 (2007).
- [44] W. J. Huang, Meng Wang, F. G. Kondev, and G. Audi, and S. Naimi, The ame 2020 atomic mass evaluation (I). Evaluation of input data and adjustment procedures, *Chinese Phys. C* **45**, 030002 (2021).
- [45] www.radiochemistry.org/periodictable/elements (2005).
- [46] S. Zhang, Y. Zhang, J. Cui, and Y. Wang, Improved semiempirical relationship for α -decay half-lives, *Phys. Rev. C* **95**, 014311 (2017).
- [47] J. P. Cui, Y. Xiao, Y. H. Gao, and Y. Z. Wang, α -decay half-lives of neutron-deficient nuclei, *Nucl. Phys. A* **987**, 99 (2019).
- [48] S. Samanta, S. Das, R. Bhattacharjee, S. Chatterjee, R. Raut, S. S. Ghugre, A. K. Sinha, U. Garg, Neelam, N. Kumar, P. Jones, Md. Sazedur R. Laskar, F. S. Babra, S. Biswas, S. Saha, P. Singh, and R. Palit, Single-particle excitations in the level structure of ^{64}Cu , *Phys. Rev. C* **97**, 014319 (2018).