Available online at www.sciencedirect.com

ScienceDirect

Nuclear Physics A 1038 (2023) 122707

NUCLEAR PHYSICS A

www.elsevier.com/locate/nuclphysa

Nuclear mass predictions with multi-hidden-layer feedforward neural network

Xian-Kai Le, Nan Wang *, Xiang Jiang *

College of Physics and Optoelectronic Engineering, Shenzhen University, 518060 Shenzhen, People's Republic of China

Received 11 April 2023; received in revised form 4 June 2023; accepted 19 June 2023
Available online 22 June 2023

Abstract

Based on Keras deep learning framework, the feedforward neural network (FNN) model is employed to improve the predictions of the liquid drop model (LDM). It is shown that the prediction ability of FNN can be significantly improved if multiple hidden layers are used with only four input parameters. The root-mean-square deviation (RMSD) of nuclear mass predicted by LDM can be reduced from 2.38 MeV to 196 keV with the multi-hidden-layer FNN model, which is only one third of the single-hidden-layer FNN model. The predictions of two-neutron separation energies (S_{2n}) and single-neutron separation energies (S_n) indicate that the multi-hidden-layer FNN model gives a better description of the shell structure. In addition, the extrapolation capability of the model in the super-heavy nuclear region is studied, the results show that better extrapolation capability can be achieved if multiple hidden layers are employed.

Keywords: Nuclear mass; Multi-hidden-layer neural network; Feedforward neural network; Liquid-drop model

1. Introduction

Accurate nuclear mass is very important in nuclear physics [1,2]. The structural information of the nucleus, such as the shell structure [3,4] and nuclear deformation [5,6] can be deduced from the nuclear mass. The neutron and proton separation energies and β decay energies can also be obtained. The study of nuclear mass is helpful to understand the existing limits of atomic

* Corresponding authors.
E-mail addresses: wangnan@szu.edu.cn (N. Wang), jiangxiang@szu.edu.cn (X. Jiang).

https://doi.org/10.1016/j.nuclphysa.2023.122707
0375-9474/

nuclei [7,8]. This is significant for the study of superheavy nuclei. In addition, nuclear mass is also very important in nuclear astrophysics [9,10], such as the r-process [11,12].

There are various nuclear mass models which can be classified as macroscopic models, macroscopic-microscopic models, and microscopic models. The root-mean-square deviation (RMSD) between the theoretical values and the experimental data is used to evaluate the prediction accuracy of the model. The macroscopic models, such as the Bethe-Weizsäcker mass formula (BW) [13,14], have a prediction accuracy of about 3 MeV for the nuclear mass. The RMSD of the macroscopic-microscopic models, for example, the finite-range droplet model (FRDM) [15] and the Weizsäcker-Skyrme (WS) model [16], is about 500 keV. The Hartree-Fock-Bogoliubov model (HFB) [17–19] is a popular microscopic model, whose prediction accuracy is similar to that of the macroscopic-microscopic models.

In recent years, neural network approaches are widely applied to improve the prediction of nuclear masses. For example, Levenberg–Marquardt neural network (LMNN) has good generalization ability in the region far away from the stability line. The RMSD of LMNN for nuclear mass prediction is 200-300 keV [20–22]. Bayesian neural network (BNN) [23–27] can effectively avoid the problem of over-fitting and can quantify the uncertainties in the predictions [28]. The RMSD of BNN can reach 200 keV [24,25], or even 100 keV [26]. Radial basis function (RBF) neural network uses hyperparameters to avoid over-fitting. It has good interpolation ability and its RMSD is 100-200 keV [29–31]. Although neural network approaches (such as BNN, RBFNN) for nuclear mass prediction have achieved good prediction results in the known nuclear region, the extrapolation results of these different neural network approaches are not consistent in the unknown region [27], which indicates that the neural network approach still has some defects.

Generally, neural networks have different hidden layers [32]. Neural networks with only one hidden layer can approximate any function that contains a continuous mapping from finite space to another. Neural networks with two hidden layers, it can represent an arbitrary decision boundary to arbitrary accuracy with rational activation functions. Meanwhile, it can approximate any smooth mapping to any accuracy. If the number of hidden layers is greater than two, this network can learn very complex features. Because there are very sophisticated underlying physics in the description of nuclear mass, the multi-hidden-layer neural network may have more advantages than the single-hidden-layer neural network.

Up to now, most of the studies on nuclear mass with the neural network have only one hidden layer, and few use multiple hidden layers [33,34]. In Ref. [33], the RMSD of the multi-hidden-layer model is higher than 1 MeV. In a recent study, 2 to 18 hidden layers are used and good results are achieved [34]. However, it can be seen that more hidden layers do not lead to smaller RMSD. It is necessary to explore the results of neural networks with different hidden layers and look for a better design of the multi-hidden-layer neural network.

In this work, we established three types of feedforward neural networks (FNN) with hidden layers of 1, 2, and 3. They are applied to improve the predictions of nuclear mass in the liquid drop model (LDM). With only four input parameters, the RMSD is reduced to 196 keV. The prediction error obtained by using similar approach is usually about 200-600 keV [20,21,34,35]. In addition, the performance of the models in predicting the single- and two-neutron separation energies and extrapolation capability in the super-heavy nuclear region is also studied. To our knowledge, this is the first time to compare the different prediction power between a single-hidden-layer neural network and a multi-hidden-layer neural network on nuclear mass.

The organization of this paper is as follows: In Sec. 2, LDM and multi-hidden-layer neural networks are introduced in detail. In Sec. 3, the predicted mass as well as the single- and two-

neutron separation energies of different models are shown, and the extrapolation capability of different models is tested in the super-heavy nuclear region. Finally, the conclusions and outlook are presented in Sec. 4.

2. Nuclear mass model

The most common way to use a machine learning approach to predict nuclear mass is to optimize a theoretical mass model. The learning objective is to minimize the difference between the experimental data B_{EXP} and the theoretical value B_{TH} of the binding energy. The output target of the neural network is:

$$\Delta B\left(Z,N\right)=B_{EXP}\left(Z,N\right)-B_{TH}\left(Z,N\right). \tag{1}$$

Our work is based on the liquid-drop model (LDM) [36], in which the binding energy can be expressed as:

$$\begin{aligned}B_{LDM}=&a_V\left(1+k_V I^2\right)A+a_S\left(1+k_S I^2\right)A^{2/3}+a_C\frac{Z^2}{A^{1/3}}\left(1-Z^{-2/3}\right)\\&+c_1\frac{2-|I|}{2+|I|\,A}I^2A+a_{pair}A^{-1/3}\delta_{np},\end{aligned} \tag{2}$$

where Z is the proton number, A is the mass number, I is equal to $(A-2Z)/A$. a_V, k_V, a_S, k_S, a_C, c_1, a_{pair} are the constants to be fitted. δ_{np} is the pair energy which can be expressed as:

$$\delta_{np}=\begin{cases}2-|I| & Z-even,\ N-even\\ |I| & Z-odd,\ N-odd\\ 1-|I| & Z-odd,\ N-even,\ N>Z\\ 1-|I| & Z-even,\ N-odd,\ N<Z\\ 1 & Z-odd,\ N-even,\ N<Z\\ 1 & Z-even,\ N-odd,\ N>Z.\end{cases} \tag{3}$$

There are seven parameters to be fitted, and the data used for fitting are selected from the AME2020 atomic mass evaluation [37]. 2429 nuclides with Z and N greater than 8 and the experimental uncertainties within 150 keV are selected from AME2020. The parameters after fitting are shown in Table 1:

Table 1
Parameters of the LDM model.

Parameter	Value
a_V	15.6606 MeV
k_V	−2.0026
a_S	−18.4157 MeV
k_S	−3.0620
a_C	−0.7149 MeV
c_1	−42.3948 MeV
a_{pair}	6.7656 MeV

The mass model is evaluated using root-mean-square deviation (RMSD):

$$\sigma=\sqrt{\frac{1}{N}\sum_n\left(B^n_{EXP}-B^n_{TH}\right)^2}. \tag{4}$$

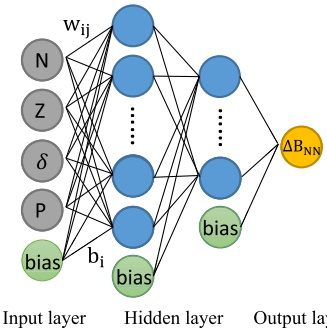

Fig. 1. Feedforward neural network with multiple hidden layers.

Because the differences between the experimental and theoretical values of mass are opposite to those of binding energies, the RMSD of binding energy calculated by Eq. (4) is also the RMSD of mass. The RMSD of the selected 2429 nuclides obtained by the LDM is 2.38 MeV.

In the early study of nuclear mass by neural network approach, there were only two input parameters (Z and A) [23]. It is found in later research that adding more physical parameters related to the nuclear structure could significantly improve the prediction results [20,25,35]. The input layer of our FNN model has four parameters: neutron number N, proton number Z, $\delta = \frac{(-1)^N+(-1)^Z}{2}$, $P = \frac{v_p v_n}{v_p+v_n}$, v_p (v_n) represents the difference between Z (N) and the nearest magic number. The output layer of the network has only one neuron, and the output is expressed as $\Delta B_{NN}(Z,N)$. The binding energy obtained by the neural network can be expressed as:

$$B_{NN}(Z,N) = B_{LDM}(Z,N) + \Delta B_{NN}(Z,N). \tag{5}$$

The detailed structure of the network is shown in Fig. 1. It takes a fully connected neural network with two hidden layers as an example to show the connection between neurons. The output of one neuron on each layer is determined by the output of the upper layer, which can be expressed as:

$$y_i = \tan\mathrm{h}\left(b_i + \sum_j w_{ij}x_j\right), \tag{6}$$

where x_j is the output value of neurons in the upper layer, w_{ij} is the weight of connections between neurons, and b_i is the bias. The activation function in the middle layer is tanh, while there is no activation function at the output layer. The optimization goal of FNN is to minimize the loss function. We use mean square deviation as the loss function of the model, which is defined as:

$$\mathrm{Loss} = \frac{1}{N}\sum_n \left(B^n_{tar} - B^n_{out}\right)^2, \tag{7}$$

where B^n_{tar} and B^n_{out} are the target value and output value of the model, respectively.

To improve the convergence performance of the model, we adopted a variable learning rate. The initial learning rate was set to 0.001. If the loss function doesn't improve within 2000 steps, the learning rate will decrease by a factor of 0.3. We take results after the fluctuation of the loss function decreases. Here, the early stopping is also used during the training to avoid the

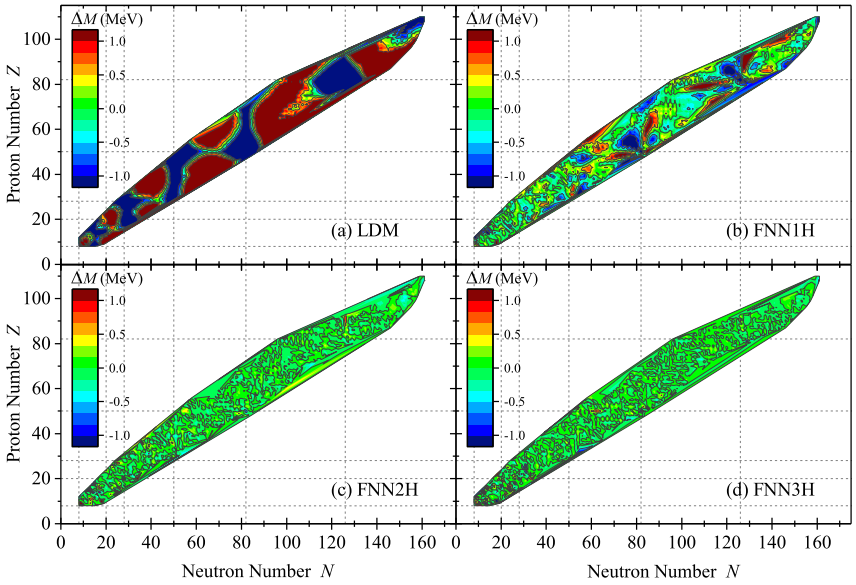

Fig. 2. Global mass prediction accuracy of (a) LDM, (b) FNN1H, (c) FNN2H, (d) FNN3H. FNNiH is used to represent FNN with i hidden layers. Magic numbers are denoted by dashed lines. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

overfitting problem. At the end of the training, the loss function tends to converge in both the learning set and the verification set.

There is no unified statement on the number of neurons in each layer. In practice there are two forms: (a) Continuous compression. The number of neurons decreases gradually in the order of input layer - hidden layer - output layer. (b) Compression after expansion. The number of neurons increases and then decreases. It is shown that the second approach may work better [38]. In this work, we use this method when designing the number of neurons in the hidden layer.

We establish three FNN models with hidden layers of 1, 2, and 3. The structure of them are 4-60-1, 4-50-25-1, 4-45-28-8-1, respectively. The number of neurons in the hidden layer is determined by the model with the smallest loss function in different designs. And the total parameters of the model with hidden layers 1, 2, and 3 are 361, 1551, and 1754, respectively.

3. Results

The experimental data in Eq. (1) are selected from AME2020. 2429 nuclides are divided into the learning set and verification set at a ratio of about 8:2. The verification set is a test for reliability. There are 1937 nuclides in the learning set and 492 in the verification set. Fig. 2 shows the difference between the experimental data and the predicted values of the LDM and FNN models. FNNiH is used to represent FNN with i hidden layers. It can be seen that the deviation of LDM is generally greater than 1 MeV. FNN1H has a great improvement, but its performance near magic numbers is not ideal. As a comparison, the prediction deviations near magic numbers are significantly reduced by using FNN2H and FNN3H, which indicates that the multi-hidden-layer FNN model has a better prediction of shell correction.

Table 2
The RMSD (keV) of different hidden layer models.

Data sets	FNN1H	FNN2H	FNN3H
Learning Set (keV)	518	117	92
Validation Set (keV)	636	370	429
Entire Set (keV)	544	196	210

Table 2 shows the calculation results of the three FNN models in detail. The RMSD of FNN1H, FNN2H, and FNN3H of the learning set are 518 keV, 117 keV, and 92 keV. While the RMSD of the verification set are 636 keV, 370 keV, and 429 keV, respectively. And the entire set gives the global mass prediction accuracy of the model, we can see that the RMSD are 544 keV, 196 keV, and 210 keV, respectively. From Table 2, we find that the prediction accuracy (2.38 MeV) of the LDM model has been significantly improved by using the FNN approach. And the results of the multi-hidden-layer FNN models are much better than those of the single-hidden-layer FNN model globally.

Note that we do not employ any complex algorithm in the FNN model, and only four input parameters are used. A global mass accuracy around 196 keV is obtained, which is excellent compared with other more complex neural network models, such as LMNN [20–22] and BNN [23–27] (the accuracy of these neural network models is around 200kev-300 keV). It indicates that the nuclear mass can be better predicted by simply increasing the depth of the neural network through appropriate parameter adjustment.

Comparing the FNN2H and FNN3H, we find that FNN2H is inferior to FNN3H in the learning set. However, in the verification set, FNN2H is better. There are only more than 3000 experimental data of nuclear mass, which limits the depth of the network. We believe that continuing to increase the depth of the network will not bring much benefit to the improvement of the model (In Ref. [34], the RMSD are not significantly improved if 2 to 18 hidden layers are used). With more physical parameters as input and more advanced network structures, whether a deeper network will yield better predictions or not is still an open question.

The nucleon separation energies are also investigated, which are important for studying the structural information (such as shell structure, pairing effects, etc.) of the nucleus. And it is also one of the criteria to evaluate the accuracy of the nuclear mass model. To further compare the differences between the three models and test the reliability of our models, the two-neutron separation energies (S_{2n}) and the single-neutron separation energies (S_n) of isotopic chains of Ni, Sn, and Pb (experimental data are selected from AME2020 [37]) are calculated and the results are shown in Fig. 3 and Fig. 4. It can be seen that although the neural network models with different hidden layers can predict the general trend of two (single)-neutron separation energies, there is a big difference between the results of the single-hidden-layer FNN model and the experimental data near the nuclides with double magic numbers ($^{78}_{28}$Ni, $^{132}_{50}$Sn, $^{208}_{82}$Pb). However, the two (single)-neutron separation energies predicted by the multi-hidden-layer FNN models are better consistent with the experimental data. In addition, the predictions of S_n show an odd-even fluctuation, which can be much improved by the multi-hidden-layer FNN models. These results not only show that the multi-hidden-layer network has more advantages than the single-hidden-layer network but also indicate that the deeper network can capture more nuclear structural information.

Finally, we test the extrapolation capability of the models for $Z = 109$, $Z = 110$, $Z = 111$, and $Z = 112$ isotopic chains which are located in the super-heavy nuclear region. Notice that

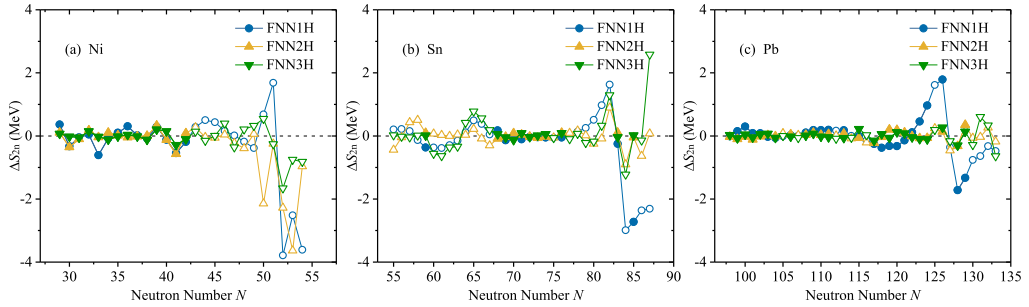

Fig. 3. Comparison of predicted two-neutron separation energies with experimental data for isotopic chains of (a) Ni, (b) Sn, (c) Pb. Solid symbols represent data in the learning region, and open symbols represent data outside the learning region.

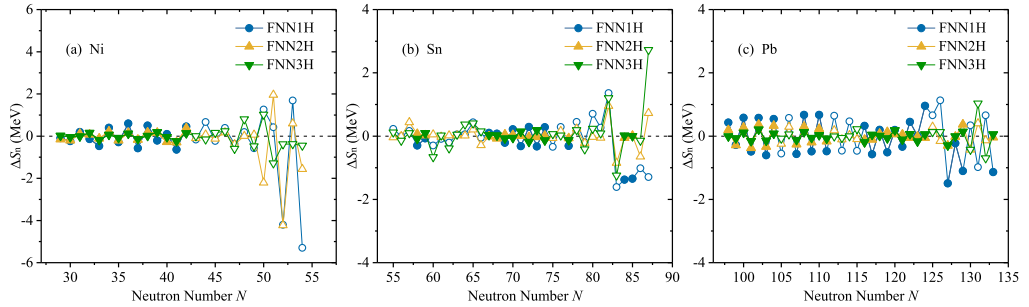

Fig. 4. Same as Fig. 3, but for single-neutron separation energies.

only $^{266}_{109}$Mt, $^{269}_{110}$Ds, $^{270}_{110}$Ds, $^{271}_{110}$Ds are selected for training, and the other nuclides' masses are not learned by the models. The predicted masses of FNNiH are presented in Fig. 5. It can be seen that the prediction power of the FNN1H decreases obviously if N is greater than 161, which is beyond the learning range. For $Z = 112$, the deviation can be as high as 10 MeV. The multi-hidden-layer FNN models, however, show an astonishing extrapolation capability, even Z and N are both outside the learning range. The maximum prediction deviation is less than 1 MeV.

4. Conclusion

FNN models with hidden layers of 1, 2, and 3 are used to improve the results of LDM. The RMSD of the LDM model decreases from 2.38 MeV to 544 keV if single-hidden-layer FNN model is applied, while it becomes 196 keV for multi-hidden-layer FNN model. This indicates that the predictive power of the neural network can be improved if multiple hidden layers are used.

In addition, the predictive power of the FNN model on the two (single)-neutron separation energies and the extrapolation capability in the super-heavy nuclear region are analyzed. The results indicate that the multi-hidden-layer neural network performs better than the single-hidden-layer neural network in all cases. Compared with single-hidden-layer neural networks, multi-hidden-layer neural networks can capture more nuclear structure information and make a better prediction.

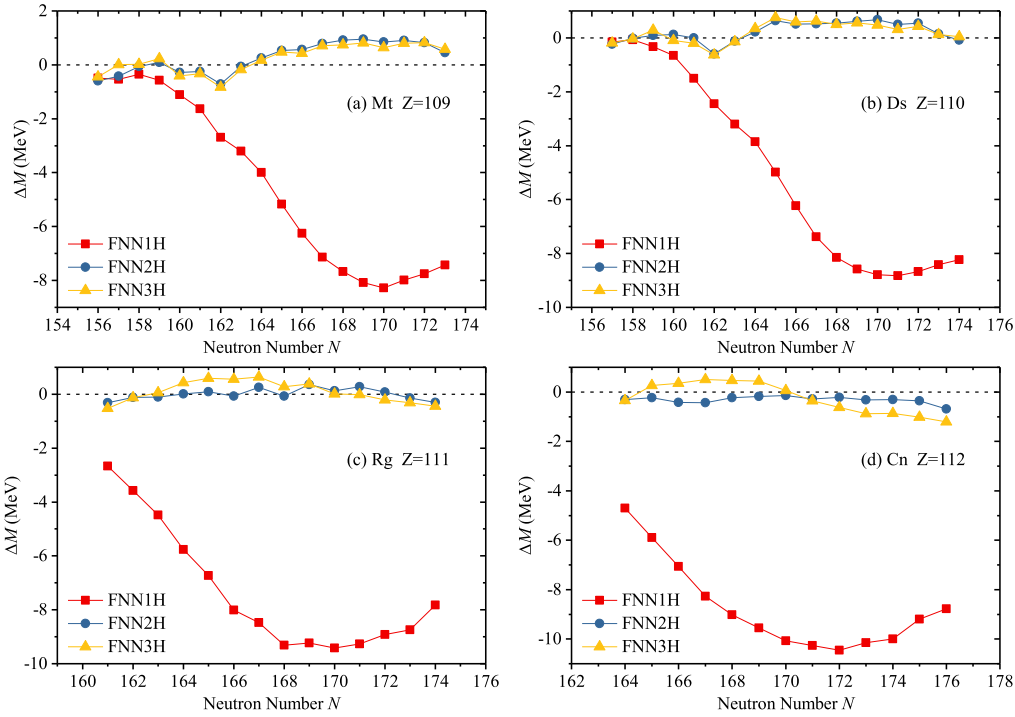

Fig. 5. Mass differences between the experimental data and the FNNiH predictions for isotopic chains of (a) $Z = 109$, (b) $Z = 110$, (c) $Z = 111$, (d) $Z = 112$.

Theoretically, the nonlinear mapping capability of a single-hidden-layer neural network is limited, and when the number of hidden layers is equal to or greater than 2, the model can approximate any smooth mapping to any precision. In this work, however, FNN3H is not better than FNN2H. As a result, more hidden layers are not always better. The most appropriate number of hidden layers should be carefully tested in practice. Over the years, the use of the neural network approach gives us a more accurate mass model, but the neural network structure we use is still fully connected. Some more advanced network structures have already been established, such as the recursive neural network [39] and self-attention neural network [40,41]. Besides introducing more physical parameters, using more advanced network structures to capture more nuclear structure information is in progress.

CRediT authorship contribution statement

Xian-Kai Le: Conceptualization, Data curation, Methodology, Software, Visualization, Writing – original draft. **Nan Wang:** Supervision, Writing – review & editing. **Xiang Jiang:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grants No. 12175151, No. 11705118, the Guangdong Major Project of Basic and Applied Basic Research under Grant No. 2021B0301030006, the Steady Support Program for Higher Education Institutions of Shenzhen under Grants No. 20200817005440001, No. 20200810163629001.

References

[1] D. Lunney, J.M. Pearson, C. Thibault, Recent trends in the determination of nuclear masses, Rev. Mod. Phys. 75 (2003) 1021.
[2] M. Bender, P.H. Heenen, P.G. Reinhard, Self-consistent mean-field models for nuclear structure, Rev. Mod. Phys. 75 (2003) 121.
[3] E.M. Ramirez, D. Ackermann, K. Blaum, M. Block, C. Droese, C.E. Düllmann, M. Dworschak, M. Eibach, S. Eliseev, E. Haettner, F. Herfurth, F.P. Heßberger, S. Hofmann, J. Ketelaer, G. Marx, M. Mazzocco, D. Nesterenko, Y.N. Novikov, W.R. Plaß, D. Rodríguez, C. Scheidenberger, L. Schweikhard, P.G. Thirolf, C. Weber, Direct mapping of nuclear shell effects in the heaviest elements, Science 337 (2012) 1207.
[4] F. Wienholtz, D. Beck, K. Blaum, C. Borgmann, M. Breitenfeldt, R.B. Cakirli, S. George, F. Herfurth, J.D. Holt, M. Kowalska, S. Kreim, D. Lunney, V. Manea, J. Menéndez, D. Neidherr, M. Rosenbusch, L. Schweikhard, A. Schwenk, J. Simonis, J. Stanja, R.N. Wolf, K. Zuber, Masses of exotic calcium isotopes pin down nuclear forces, Nature 498 (2013) 346.
[5] U. Hager, T. Eronen, J. Hakala, A. Jokinen, V.S. Kolhinen, S. Kopecky, I. Moore, A. Nieminen, M. Oinonen, S. Rinta-Antila, J. Szerypo, J. Äystö, First precision mass measurements of refractory fission fragments, Phys. Rev. Lett. 96 (2006) 042504.
[6] A. de Roubin, D. Atanasov, K. Blaum, S. George, F. Herfurth, D. Kisler, M. Kowalska, S. Kreim, D. Lunney, V. Manea, E. Minaya Ramirez, M. Mougeot, D. Neidherr, M. Rosenbusch, L. Schweikhard, A. Welker, F. Wienholtz, R.N. Wolf, K. Zuber, Nuclear deformation in the a $\approx$ 100 region: comparison between new masses and mean-field predictions, Phys. Rev. C 96 (2017) 014310.
[7] K. Rutz, M. Bender, T. Bürvenich, T. Schilling, P.G. Reinhard, J.A. Maruhn, W. Greiner, Superheavy nuclei in self-consistent nuclear calculations, Phys. Rev. C 56 (1997) 238.
[8] M. Liu, N. Wang, Y. Deng, X. Wu, Further improvements on a global nuclear mass model, Phys. Rev. C 84 (2011) 014333.
[9] H.A. Bethe, Energy production in stars, Phys. Rev. 55 (1939) 434.
[10] E.M. Burbidge, G.R. Burbidge, W.A. Fowler, F. Hoyle, Synthesis of the elements in stars, Rev. Mod. Phys. 29 (1957) 547.
[11] M. Mumpower, R. Surman, G. McLaughlin, A. Aprahamian, The impact of individual nuclear properties on r-process nucleosynthesis, Prog. Part. Nucl. Phys. 86 (2016) 86.
[12] D. Martin, A. Arcones, W. Nazarewicz, E. Olsen, Impact of nuclear mass uncertainties on the r-process, Phys. Rev. Lett. 116 (2016) 121101.
[13] C.F.v. Weizsäcker, Zur theorie der kernmassen, Z. Phys. 96 (1935) 431.
[14] H.A. Bethe, R.F. Bacher, Nuclear Physics A, Stationary states of nuclei, Rev. Mod. Phys. 8 (1936) 82.
[15] P. Möller, A. Sierk, T. Ichikawa, H. Sagawa, Nuclear ground-state masses and deformations: FRDM(2012), At. Data Nucl. Data Tables 109 (2016) 1.
[16] N. Wang, M. Liu, X.Z. Wu, J. Meng, Surface diffuseness correction in global mass formula, Phys. Lett. B 734 (2014) 215.
[17] S. Goriely, N. Chamel, 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 (2009) 152503.
[18] S. Goriely, N. Chamel, J.M. Pearson, Further explorations of Skyrme-Hartree-Fock-Bogoliubov mass formulas, XVI, Inclusion of self-energy effects in pairing, Phys. Rev. C 93 (2016) 034337.

[19] S. Goriely, S. Hilaire, M. Girod, S. Péru, First Gogny-Hartree-Fock-Bogoliubov nuclear mass model, Phys. Rev. Lett. 102 (2009) 242501.
[20] T.L. Zhao, H.F. Zhang, A neural network approach based on more input neurons to predict nuclear mass, Chin. Phys. C 46 (2022) 044103.
[21] T.L. Zhao, H.F. Zhang, A new method to improve the generalization ability of neural networks: a case study of nuclear mass training, Nucl. Phys. A 1021 (2022) 122420.
[22] H.F. Zhang, L.H. Wang, J.P. Yin, P.H. Chen, H.F. Zhang, Performance of the Levenberg-Marquardt neural network approach in nuclear mass prediction, J. Phys. G, Nucl. Part. Phys. 44 (2017) 045110.
[23] R. Utama, J. Piekarewicz, H.B. Prosper, Nuclear mass predictions for the crustal composition of neutron stars: a Bayesian neural network approach, Phys. Rev. C 93 (2016) 014311.
[24] R. Utama, J. Piekarewicz, Validating neural-network refinements of nuclear mass models, Phys. Rev. C 97 (2018) 014306.
[25] Z.M. Niu, H.Z. Liang, Nuclear mass predictions based on Bayesian neural network approach with pairing and shell effects, Phys. Lett. B 778 (2018) 48.
[26] Z.M. Niu, H.Z. Liang, Nuclear mass predictions with machine learning reaching the accuracy required by r-process studies, Phys. Rev. C 106 (2022) L021303.
[27] Z.M. Niu, J.Y. Fang, Y.F. Niu, Comparative study of radial basis function and Bayesian neural network approaches in nuclear mass predictions, Phys. Rev. C 100 (2019) 054311.
[28] J. Lampinen, A. Vehtari, Bayesian approach for neural networks—review and case studies, Neural Netw. 14 (2001) 257–274.
[29] N. Wang, M. Liu, Nuclear mass predictions with a radial basis function approach, Phys. Rev. C 84 (2011) 051303(R).
[30] J.S. Zheng, N.Y. Wang, Z.Y. Wang, Z.M. Niu, Y.F. Niu, B. Sun, Mass predictions of the relativistic mean-field model with the radial basis function approach, Phys. Rev. C 90 (2014) 014303.
[31] N.N. Ma, H.F. Zhang, X.J. Bao, Basic characteristics of nuclear landscape by improved Weizsäcker-Skyrme-type nuclear mass model, Chin. Phys. C 43 (2019) 044105.
[32] J. Heaton, Heaton research: the number of hidden layers, https://www.heatonresearch.com/2017/06/01/hidden-layers, 2017.
[33] E. Yüksel, D. Soydaner, H. Bahtiyar, Nuclear binding energy predictions using neural networks: application of the multilayer perceptron, Int. J. Mod. Phys. E 30 (2021) 2150017.
[34] C.Q. Li, C.N. Tong, H.J. Du, Deep learning approach to nuclear masses and α-decay half-lives, Phys. Rev. C 105 (2022) 064306.
[35] T.L. Zhao, H.F. Zhang, Neural network approach to improve the quality of atomic nuclei, Sci. China, Phys. Mech. Astron. 52 (2022) 74.
[36] N. Wang, Z. Liang, M. Liu, X.Z. Wu, Mirror nuclei constraint in nuclear mass formula, Phys. Rev. C 82 (2010) 044304.
[37] M. Wang, W.J. Huang, F.G. Kondev, G. Audi, S. Naimi, The AME 2020 atomic mass evaluation (II). Tables, graphs, and references, Chin. Phys. C 45 (2021) 030003.
[38] M. Noshad, Y. Zeng, A.O. Hero, Scalable Mutual Information Estimation Using Dependence Graphs, ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, p. 2962.
[39] J.B. Pollack, Recursive distributed representations, Artif. Intell. 46 (1990) 77.
[40] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need. Advances in neural information processing systems, https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf, 2017.
[41] P. Shaw, J. Uszkoreit, A. Vaswani, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Self-attention with relative position representations, arXiv:1803.02155.