Optimizing multilayer Bayesian neural networks for evaluation of fission yields

Zi-Ao Wang and Junchen Pei[*]

State Key Laboratory of Nuclear Physics and Technology, School of Physics, Peking University, Beijing 100871, China

Bayesian machine learning is a promising tool for the evaluation of nuclear fission data but its potential capability has not been fully realized. We attempt to optimize the performances of the multilayer Bayesian neural networks for evaluations of fission yields. The influences of adjustments of learning data, activation functions, and network structures have been studied. In particular, negative values of net functions have been penalized to avoid nonphysical inferences of fission yields. Presently the network with double hidden layers has optimal performances compared to the single-layer or deeper networks. These studies are essential for further developments of precise evaluation methods.

I. INTRODUCTION

Nuclear data evaluation is a crucial process that connects nuclear data, nuclear modelings, and nuclear applications. In particular, nuclear fission data is the key ingredient in many nuclear applications [1]. For example, high-precision and reliable neutron-induced fission product yield (FPY) distributions of actinides are very valuable. However, experimental measurements of FPY with continuous incident neutron energies are extremely difficult and insufficient. In major nuclear data libraries (ENDF [2], JENDL [3], JEFF [4], CENDL [5], etc.), complete evaluations of FPY are only available for neutron incident energies around thermal energies, 0.5 MeV and 14 MeV. There are some incomplete experimental FPY at other energies with large uncertainties. On the other hand, a predictive theoretical description of fission observables is still very challenging [6]. There are various fission models that have different advantages and applicabilities [7–13]. Microscopic fission models are promising but are not ready yet for accurate quantitative applications. Therefore, the prediction and evaluation of energy-dependent FPY for fast reactors are very anticipated.

Machine learning is a very powerful tool for learning and inference from complex big data. In recent years, machine learning has attracted great interest in various physics disciplines. Recently, it was shown that Bayesian neural network (BNN) can be used for evaluations of incomplete fission yields with uncertainty quantifications [14]. Machine learning has been used in nuclear physics with increasing interests [15], such as the extrapolation of nuclear masses [16–18], fission yields [14,19], various nuclear structures [20–25], and reaction observables [26–28]. Machine learning has also been widely applied in other physics subjects, such as the constraints of equation of state of neutron stars from gravitational wave signals [29] and for facilitating the lattice QCD calculations [30]. Conventionally, the evaluation of fission yields is based on the least-squares adjustments of parameters of various phenomenological models [31], such as the Brosa model and the GEF model [32,33]. These evaluations could not be applicable when very few experimental data points are available. Machine learning is promising for developing new evaluation methods of nuclear data, regarding the handling of various correlated fission observables with large discrepancies and uncertainties.

Previously we have demonstrated that BNN can be used for evaluation of incomplete fission mass yields [14] and fission charge yields [34]. In this work, we aim to improve the performances of multilayer Bayesian neural networks for precise evaluations of fission yields. The potential capability of BNN evaluation has not been fully realized. In principle, the BNN approach can be optimized for specific applications. For example, the learning performance and overfitting are two competing issues in machine learning. With more layers and neurons in the network, its learning performance increases while overfitting issues also appear, which undermines its predictive capability. Therefore, more studies about the sensitivity of configurations of neural networks are needed. Recently, the physics-informed machine learning is a hot topic [35–38], which could be very useful for nuclear physics when available observational data are not sufficient. The physics-informed machine learning can be implemented through the physics-informed prior [35], or physics-constrained cost function [36], or physics-augmented input variables [18,34], or physics-inspired network architectures [38]. Previously we reported that BNN can result in unphysical negative values of fission yields [14]. It is interesting to build physics constraints into BNN to avoid negative yields. In this work, we plan to study the influences of adjustments of learning data, the choice of activation functions, the structure of multilayer networks, and the penalty of negative values to constrain the fission yields. Finally, as an example, we demonstrate that BNN is used for precise evaluations of fission mass yields of ^{239}U.
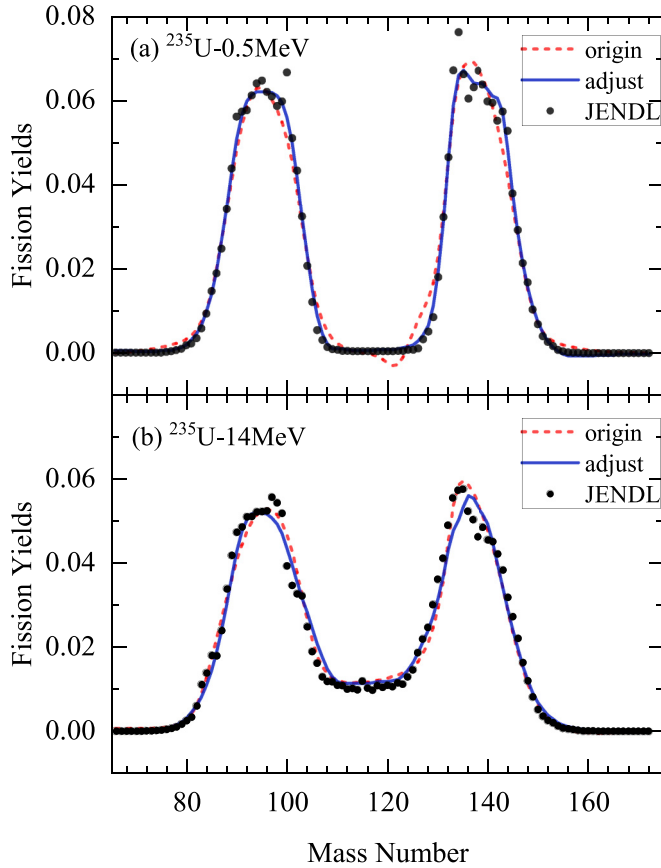
_____
*peij@pku.edu.cn

FIG. 1. The BNN learning performances are shown for neutron-induced fission yields of ^{235}U at 0.5 and 14 MeV, respectively. The results obtained by learning the original fission data or the adjusted data are compared. The adjustment of data denotes a linear transformation so that the ranges of data values are within $(-0.9, 0.90)$. The learning target is taken from the evaluated data of JENDL [3].

II. THEORETICAL FRAMEWORK

The BNN approach [39] to statistical regression inference is based on Bayes' theorem, which provides a connection between a given hypothesis(in terms of problem-specific beliefs for a set of parameters) and a set of data to a posterior probability that is used to make predictions on new inputs. The BNN approach adopts probability distributions as connection weights and is naturally suitable for uncertainty quantifications, in contrast to standard neural networks, which optimize definite values for connection weights. The basic BNN is written as,

$$p(\omega|x, t) = \frac{p(x, t|\omega)p(\omega)}{p(x, t)}, \tag{1}$$

where ω denotes the network parameters, including network weights and bias; $p(x, t \mid \omega)$ is the likelihood that a given model describes the data, and $p(\omega)$ is the prior distribution of the parameters ω; x and t are input and output data; $p(\omega \mid x, t)$ is the posterior distribution, i.e., the probability distribution of parameters ω after considering the data (x, t); $p(x, t)$ is a normalization factor, which ensures the integral of posterior distribution is one.

We adopt a Gaussian distribution for the likelihood based on a cost function, which is written as

$$p(x, t \mid \omega) = \exp(-\chi^2/2), \tag{2}$$

where the cost function $\chi^2(\omega)$ reads:

$$\chi^2(\omega) = \sum_{i=1}^{N} \left(\frac{t_i - f(x_i, \omega)}{\Delta t_i} \right)^2, \tag{3}$$

Here N is the number of data points, and Δt_i is the associated noise scale, which is related to specific observables. The net function $f(x_i, \omega)$ depends on the input data x_i and the model parameters ω. In this work, the inputs of the network are given by $x_i = \{A_{fi}, Z_i, A_i, E_i\}$, which include the mass number A_{fi} of the fission fragments, the charge number Z_i and mass number A_i of the fission nuclei, and the excitation energy of the compound nucleus $E_i = e_i + S_i$ (e_i and S_i are incident neutron energy and neutron separation energy, respectively); t_i are the fission mass yields. For the evaluation of fission yields, it is crucial to learn the yields individually and it is difficult to learn complete distributions as a target. Fortunately, the normalization of fission yields is kept within an uncertainty of 2%.

The posterior distributions are obtained by learning the given data. With new data x_n, we make predictions by integrating the neural network over the posterior probability density of the network parameters ω,

$$\langle f_n \rangle = \int f(x_n, \omega) p(\omega \mid x, t) d\omega. \tag{4}$$

The high-dimensional integral in Eq. (4) is approximated by Monte Carlo integration in which the posterior probability $p(\omega \mid x, t)$ is sampled using the Markov chain Monte Carlo method [39].

In BNN we need to specify the form of the functions $f(x, \omega)$ and $p(\omega)$. In this work, we use a feed-forward neural network model defined the function $f(x, \omega)$. That is

$$f(x, \omega) = a + \sum_{j=1}^{H} b_j \tanh \left(c_j + \sum_{i=1}^{I} d_{ji} x_i \right), \tag{5}$$

where H is the number of neurons in the hidden layer, I denotes the number of input variables, and $\omega = \{a, b_j, c_j, d_{ji}\}$ is the model parameters, a is the bias of output layers, b_j are the weights of output layers, c_j is the bias of hidden layers, and d_{ji} are weights of hidden layers. In total, the number of parameters in this neural network is $1 + (2 + I) \times H$. We adopt the commonly used tanh as the activation function and other nonlinear activation functions have also been tested in this work. The confidence interval (CI) at 95% level is given for uncertainty quantifications in this work. More details about BNN can be found in Ref. [39].

III. RESULTS AND DISCUSSIONS

First we study the influences of adjustments of the input data. In principle, the network can take any data without range limits. However, we speculate that the network is more sensitive to a specific range of data. For example, the tanh
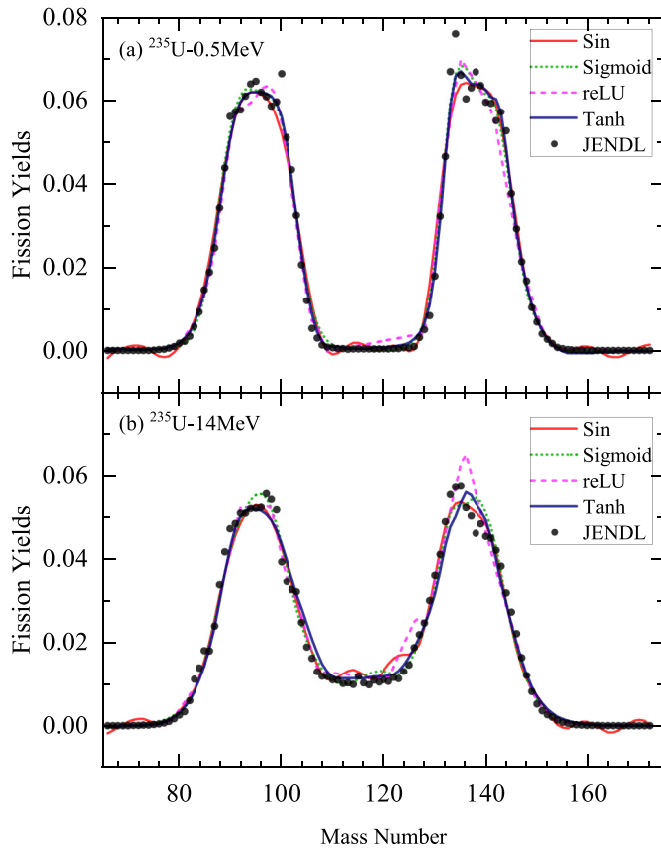
FIG. 2. The learning performances with different activation functions (sine, tanh, sigmoid, ReLU) are shown for fission yields of $n + {}^{235}$U at 0.5 and 14 MeV, respectively.

FIG. 3. The BNN evaluation of incomplete neutron-induced fission yields of (a) ^{235}U at 3.6 MeV, and (b) ^{238}U at 4.49 MeV. The normal BNN evaluation used one hidden layer with 40 neurons. The BNN results with learning penalty on negative values are shown for comparison, denoted as 40n. The experimental data is taken from [41].

activation function is not sensitive to large values and the response is within the saturation range. Actually, the tanh function is more sensitive from -0.9 to 0.9. To this end, we performed linear transformations of the original data so that both input and output data are within the range of (-0.9, 0.9). Present calculations employed 5350 evaluated fission yield data of 30 nuclei from JENDL, which is similar to our previous work [14]. Figure 1 displays the learning results of $n + {}^{235}$U at 0.5 MeV and 14 MeV, respectively. Note that calculations in Figs. 1–3 employed one hidden layer with 40 neurons. It can be seen that the learning performance with the original data is not satisfactory at 0.5 MeV. On the other hand, the learning results with adjusted data are rather good without unphysical oscillations. At 14 MeV, the two results are comparable. Previous studies also show that machine learning is difficult to describe the fission yields at low excitation energies with more quantum effects [34]. This points out that the adjustment of ranges of input data indeed has advantages for BNN performances. The adjusted normalization of input data is adopted in following studies in this work.

In neural networks, it is crucial to use nonlinear activation functions to compute nontrivial problems. There are various activation functions been adopted in neural networks [40] and it is interesting to choose a particular activation function for specific problems. In this work, we did testing calculations with tanh function, sigmoid function, ReLU (rectified linear
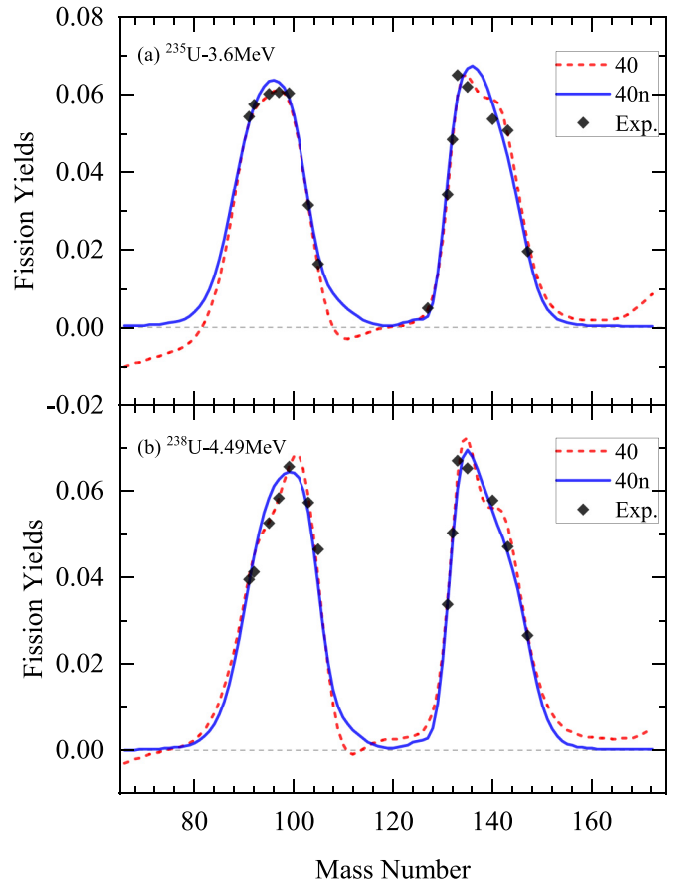
unit) function, and sine function, as shown in Fig. 2. We see that all activation functions can largely reproduce the fission yields. However, sine and ReLU functions result in some unphysical oscillations. The results of tanh are slightly better than that of sigmoid. With 5350 points, the total $\chi_N^2 = \sum_i [t_i - f(x_i)]^2 / N$ for tanh, sigmoid, ReLU, and sine activation functions are 5.78×10^{-6}, 6.37×10^{-6}, 8.79×10^{-6}, and 7.56×10^{-6}, respectively. Note that the total learning steps are 10^5 for all activation functions. We also see that sine and tanh functions can have unphysical negative fission yields, which are suppressed in sigmoid and ReLU functions. Generally, we see that tanh activation function has the best learning performance, regarding that not only it has the smallest χ_N^2 but also its descriptions of overall shapes of fission yields.

In evaluation of fission yields, the physical output values should always be positive. However, the output ranges of tanh function could be negative. This is a serious problem for inferences when data is sparse [14]. To solve this problem, we add penalty to constrain the fission yields. Actually, the weights of the likelihood function are increased [e.g., reduce the noise scale in Eq. (3)] when negative values appear in the
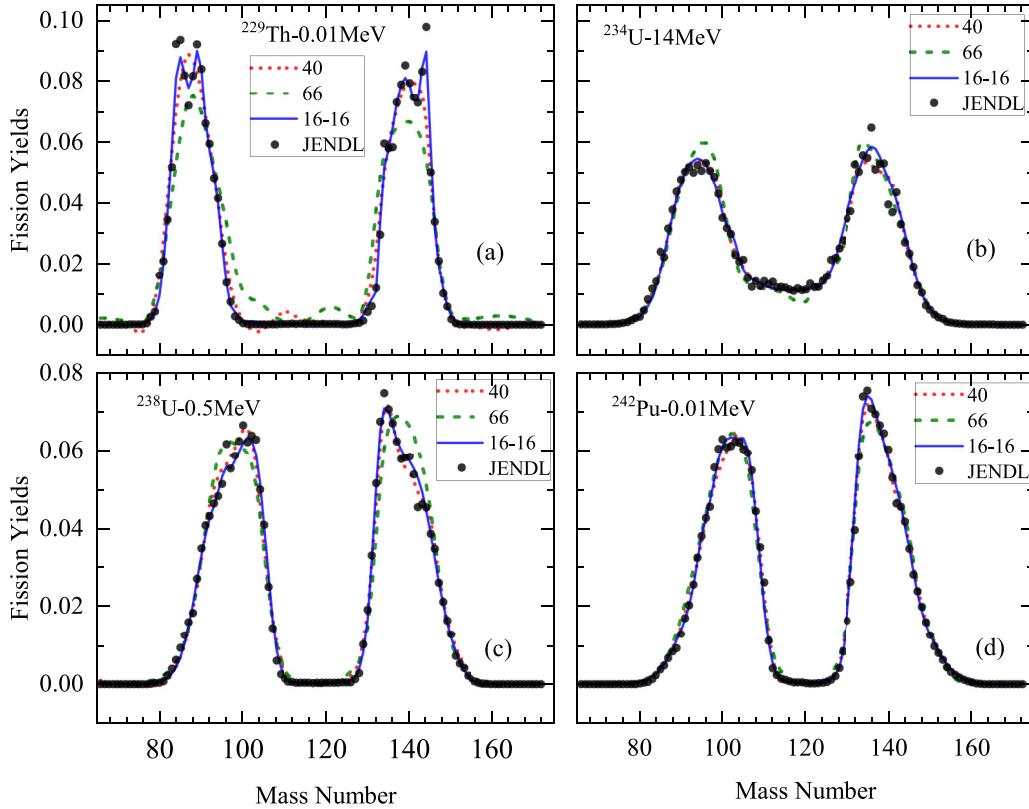
FIG. 4. BNN learning performances are shown with different networks: one layer with 40 neurons, one layer with 66 neurons, two layers with 16-16 neurons.

learning process. Figure 3 shows that evaluation of incomplete fission yields from neutron included fission of ^{235}U at 3.6 MeV and ^{238}U at 4.49 MeV. We see that BNN evaluation without penalty leads to serious unphysical negative values. The results obtained with penalized learning have much better performance when fission yields are close to zero. As a compromise, we also see that the data points can be better reproduced by BNN without penalty. Nevertheless, the penalized learning is necessary to avoid the unphysical negative values. This is also a successful attempt to build physical constraints into machine learning, towards a physics-informed machine learning.

Next we explored the optimal structure of neural networks for evaluation of fission yields. Figure 4 shows the learning results of one hidden layer with 40 neurons and 66 neurons, and two hidden layers with 16-16 neurons. We can see that results of one layer with 40 neurons are largely satisfactory. The double layers with 16-16 neurons has the best performance. The learning performances of one layer with 66 neurons are not so good with some oscillations. The total χ_N^2 of 40, 66, and 16-16 networks are 5.78×10^{-6}, 4.35×10^{-6}, and 3.43×10^{-6}, respectively. The average performance of one layer of 66 neurons is better than that of 40 neurons. But the performances of 66 neurons are not always better than that of 40 neurons. Note that the number of connection weights of the 16-16 structure is close to that of the one layer of 66 neurons, but the performance of the double-layer structure is much better. This shows the limitation of the one hidden layer structure for complex data.

Generally, with similar number of connection weights, the shallow network would be more dependent on the prior input. In contrast, the deep network would be more dependent on its deduction capability. Therefore, for specific problems, there should be a balanced choice of network structure. We also tested network structures of 11-12-12, 9-10-10-10, 9-9-8-8-9, 8-8-8-8-7-7, 7-7-7-7-7-7-8 neurons for 3, 4, 5, 6, 7 hidden layers, respectively. Note that all these structures have similar number of connection weights to that of the 16-16 structure. Correspondingly, the total χ_N^2 are 4.07×10^{-6} (three layers), 4.99×10^{-6} (four layers), 4.64×10^{-6} (five layers), 5.05×10^{-6} (six layers), 4.94×10^{-6} (seven layers). We see that deep networks have no advantages in this work. The best network is the double-layer structure of 16-16 neurons. In addition, the deep networks take much longer computing time to get convergence.

Finally, we performed BNN evaluations of fission mass yields of ^{239}U with the double-layer network, as shown in Fig. 5. The incomplete experimental data is taken from Ref. [42], in which the compound nucleus ^{239}U is produced through one neutron transfer reaction of ^{238}U + ^{9}Be. The excitation energy of ^{239}U is about 8.3 MeV [42]. The recent experiments can obtain precise isotopic identification of some fragment isotopes. For comparison, BNN evaluations with and without negative penalty are given. We see that both evaluations are satisfactory with some slight discrepancies. It has to be pointed out that in the first peak, both evaluations underestimate the fission yields. In this respect, our evaluations are consistent with the GEF evaluation. For the
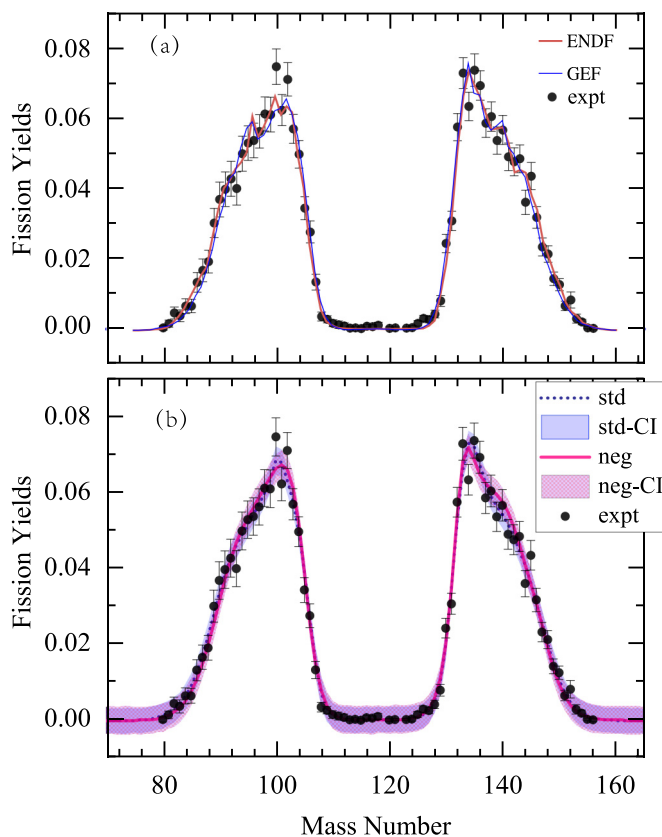
FIG. 5. The evaluations of fission yields of compound nucleus ^{239}U are shown. The experimental data is taken from [42]. GEF and ENDF evaluations are shown in (a), which are taken from [42]. The double-layer BNN evaluations without (denoted as std) and with penalty on negative values (denoted as neg) are shown in (b). The shadows in (b) denote the BNN confidence interval at 95%.

second peak, the largest value is at $A = 134$ in BNN with negative penalty, which is consistent with GEF and ENDF evaluations, in contrast to the experiment. The fission yields at $A = 140$ with negative penalty is slightly larger than that of the standard BNN evaluation, which is consistent with GEF and ENDF evaluations. Generally, the shape of the second peak obtained with negative penalty is more consistent with

GEF and ENDF evaluations, compared to the standard BNN evaluation. We demonstrated that the BNN evaluation with negative penalty is essential to describe detailed peak structures and obtain high-precision evaluations. The confidence interval from BNN with negative penalty is slightly larger than that from standard BNN due to additional constraints. Within confidence intervals, it is difficult to distinguish these two evaluations. Note that the uncertainty propagation and quantification are very important in evaluations [43], which will be studied in detail in a forthcoming work.

IV. SUMMARY

In summary, we studied the multilayer Bayesian neural networks to improve their performance for evaluations of fission yields. We investigated the influences of adjustments of ranges of input and output data for neural networks. It is useful to use a linear transformation to prepare the learning data set within a proper range so that the response saturation range of the active function can be avoided. We also studied various active functions and found that tanh function has the best learning performance. In the evaluation of fission yields, there is a serious problem that unphysical negative values can appear in outputs of neural networks. We deployed the penalized function in likelihood function to constrain the outputs. Thus negative values can be much suppressed. This is a successful attempt to implement physics constraints in neural networks. We also studied various network structures from a single hidden layer to seven hidden layers. We found that the double-layer network is optimal for the present work. Finally we performed the BNN evaluation of fission yields of ^{239}U. The results show that BNN with negative penalty is essential to obtain detailed peak structures and high-precision evaluations. This work demonstrated that BNN is a promising tool for high-precision evaluations of fission data and work in this direction is in progress.

[1] L. A. Bernstein, D. A. Brown, A. J. Koning, B. T. Rearden, C. E. Romano, A. A. Sonzogni, A. S. Voyles, and W. Younes, Ann. Rev. Nucl. Part. Sci. **69**, 109 (2019).

[2] M. B. Chadwick *et al.*, Nucl. Data Sheets **112**, 2887 (2011).

[3] K. Shibata *et al.*, J. Nucl. Sci. Tech. **48**, 1 (2011).

[4] Joint Evaluated Fission and Fusion (JEFF) Nuclear Data Library, https://www.oecd-nea.org/dbdata/jeff/.

[5] Z. G. Ge, Z. X. Zhao, H. H. Xia, Y. X. Zhuang, T. J. Liu, J. S. Zhang, and H. C. Wu, J. Korean Phys. Soc. **59**, 1052 (2011).

[6] M. Bender *et al.*, J. Phys. G **47**, 113002 (2020).

[7] N. Schunck and L. M. Robledo, Rep. Prog. Phys. **79**, 116301 (2016).

[8] K.-H. Schmidt and B. Jurado, Rep. Prog. Phys. **81**, 106301 (2018).

[9] Y. Qiang, J. C. Pei, and P. D. Stevenson, Phys. Rev. C **103**, L031304 (2021).

[10] A. Bulgac, S. Jin, K. J. Roche, N. Schunck, and I. Stetcu, Phys. Rev. C **100**, 034615 (2019).

[11] M. R. Mumpower, P. Jaffke, M. Verriere, and J. Randrup, Phys. Rev. C **101**, 054607 (2020).

[12] L. L. Liu, X. Z. Wu, Y. J. Chen, C. W. Shen, Z. X. Li, and Z. G. Ge, Phys. Rev. C **99**, 044614 (2019).

[13] J. Sadhukhan, S. A. Giuliani, Z. Matheson, and W. Nazarewicz, Phys. Rev. C **101**, 065803 (2020).

[14] Z. A. Wang, J. C. Pei, Y. Liu, and Y. Qiang, Phys. Rev. Lett. **123**, 122501 (2019).

[15] P. Bedaque, A. Boehnlein, M. Cromaz *et al.*, Eur. Phys. J. A **57**, 100 (2021).

[16] R. Utama, J. Piekarewicz, and H. B. Prosper, Phys. Rev. C **93**, 014311 (2016).

[17] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, Phys. Rev. C **98**, 034318 (2018); L. Neufcourt, Y. Cao, W. Nazarewicz, E. Olsen, and F. Viens, Phys. Rev. Lett. **122**, 062502 (2019).

[18] Z. M. Niu and H. Z. Liang, Phys. Lett. B **778**, 48 (2018).

[19] A. E. Lovell, A. T. Mohan, and P. Talou, J. Phys. G **47**, 114001 (2020).

[20] Z. M. Niu, H. Z. Liang, B. H. Sun, W. H. Long, and Y. F. Niu, Phys. Rev. C **99**, 064307 (2019).

[21] R. Utama, W. C. Chen, and J. Piekarewicz, J. Phys. G **43**, 114002 (2016).

[22] Di Wu, C. L. Bai, H. Sagawa, and H. Q. Zhang, Phys. Rev. C **102**, 054323 (2020).

[23] J. W. T. Keeble and A. Rios, Phys. Lett. B **809**, 135743 (2020).

[24] R.-D. Lasseri, D. Regnier, J.-P. Ebran, and A. Penon, Phys. Rev. Lett. **124**, 162502 (2020).

[25] W. G. Jiang, G. Hagen, and T. Papenbrock, Phys. Rev. C **100**, 054326 (2019).

[26] C. W. Ma, D. Peng, H. L. Wei, Z. M. Niu, Y. T. Wang, and R. Wada, Chin. Phys. C **44**, 014104 (2020).

[27] C. W. Ma, D. Peng, H. L. Wei, Y. T. Wang, and J. Pu, Chin. Phys.C **44**, 124107 (2020).

[28] A. E. Lovell, F. M. Nunes, M. Catacora-Rios, and G. B. King, J. Phys. G **48**, 014001 (2021).

[29] B. P. Abbott *et al.*, Phys. Rev. Lett. **121**, 161101 (2018).

[30] P. E. Shanahan, D. Trewartha, and W. Detmold, Phys. Rev. D **97**, 094506 (2018).

[31] T. R. England, B. F. Rider, Evaluation and compilation of fission product yields, LA-UR-94-3106 (Los Alamos National Laboratory, 1993).

[32] U. Brosa, S. Grossmann, and A. Müller, Phys. Rep. **197**, 167 (1990).

[33] K.-H. Schmidt, B. Jurado, C. Amouroux, and C. Schmitt, Nucl. Data Sheets **131**, 107 (2016).

[34] C. Y. Qiao, J. C. Pei, Z. A. Wang, Y. Qiang, Y. J. Chen, N. C. Shu, and Z. G. Ge, Phys. Rev. C **103**, 034621 (2021).

[35] D. R. Phillips, R. J. Furnstahl, U. Heinz, T. Maiti, W. Nazarewicz, F. M. Nunes, M. Plumlee, M. T. Pratola, S. Pratt, F. G. Viens, and S. M. Wild, J. Phys. G **48**, 072001 (2021).

[36] X. Jia, J. Willard, A. Karpatne, J. S. Read, J. A. Zwart, M. Steinbach, V. Kumar, ACM/IMS Trans. Data Sci. **2**, 1 (2021).

[37] D. Neudecker, O. Cabellos, A. R. Clark, M. J. Grosskopf, W. Haeck, M. W. Herman, J. Hutchinson, T. Kawano, A. E. Lovell, I. Stetcu, P. Talou, and S. V. Wiel, Phys. Rev. C **104**, 034611 (2021).

[38] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, Nature Rev. Phys. **3**, 422 (2021).

[39] R. Neal, *Bayesian Learning of Neural Network* (Springer, New York, 1996).

[40] https://en.wikipedia.org/wiki/Activation_function.

[41] M. E. Gooden *et al.*, Nucl. Data Sheets **131**, 319 (2016).

[42] D. Ramos, M. Caamano, A. Lemasson, M. Rejmund, H. Alvarez-Pol, L. Audouin, J. D. Frankland, B. Fernandez-Dominguez, E. Galiana-Baldo, J. Piot, C. Schmitt, D. Ackermann, S. Biswas, E. Clement, D. Durand, F. Farget, M. O. Fregeau, D. Galaviz, A. Heinz, A. Henriques, B. Jacquot, B. Jurado, Y.H. Kim, P. Morfouace, D. Ralet, T. Roger, P. Teubig, and I. Tsekhanovich, Phys. Rev. C **101**, 034609 (2020).

[43] P. Talou, P. G. Young, T. Kawano, M. Rising, and M. B. Chadwick, Nucl. Data Sheets **112**, 3054 (2011).