Improved naive Bayesian probability classifier in predictions of nuclear mass

Yifan Liu,[1] Chen Su,[1] Jian Liu,[1,2,*] Pawel Danielewicz,[2] Chang Xu,[3] and Zhongzhou Ren[4]
[1]*College of Science, China University of Petroleum (East China), Qingdao 266580, China*
[2]*National Superconducting Cyclotron Laboratory and Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA*
[3]*Department of Physics, Nanjing University, Nanjing 210093, China*
[4]*School of Physics Science and Engineering, Tongji University, Shanghai 200092, China*

(Received 31 August 2020; accepted 6 July 2021; published 16 July 2021)

Background: Recently, novel statistical methods such as neural networks and Bayesian learning methods are implemented to describe the nuclear masses.
Purpose: Based on previous studies, an improved naive Bayesian probability (iNBP) classifier is proposed to study the nuclear masses by refining the results of sophisticated nuclear models.
Method: In the iNBP method, the prediction for nuclear masses is treated as a classification problem. The residuals are classified into several groups to generate prior and conditional probabilities, and the posterior probabilities are further determined by the Bayesian formula. We choose the expectation with maximum probability as the final prediction. Reliability of the iNBP method is assessed by analyzing the global optimizations and the extrapolating capabilities.
Results: The iNBP method exhibits impressive improvements on global descriptions for different mass models. Moreover, the method shows robust extrapolating capabilities. Results demonstrate the iNBP method can be applied to predict the nuclear masses of unknown regions.
Conclusions: Considering the local mass relations, the iNBP method can offer considerable fine-tuning of the mass descriptions from nuclear models. The methodology proposed in this paper can also be applied to other model-based extrapolations of nuclear observables.

DOI: 10.1103/PhysRevC.104.014315

I. INTRODUCTION

The mass of a nucleus, originating from complex interactions between nucleons, is a basic and indispensable property in nuclear physics. Researches on nuclear masses not only provide guidance in searches for superheavy nuclei [1,2], but also help understand problems in nuclear structure [3–5], nuclear decays [6–8], nuclear reactions [9–11], and nuclear astrophysics [12–14]. There are mainly two types of experimental methods to measure the nuclear masses. The direct methods are based on mass spectrometry, such as Penning trap [15] and storage ring [16]; while the indirect methods constrain nuclear masses by determining the Q values of nuclear reactions or decays [17,18].

As more and more nuclear mass data are being filled up in the mass database, there has been a great development in theoretical mass models. Three classes of nuclear models are proposed to describe changing rules of nuclear mass. The first class of these models are the local mass models such as the Garvey-Kelson relations (GKs) [19–22], which can predict unknown masses from neighboring nuclei. The predictions of the local mass models are typically accurate, but the poor behaviors in extrapolation become a limitation of such models [20]. Secondly, there are some macroscopic models including the finite-range droplet model (FRDM) [23] and other semiempirical models with microscopic corrections [24–26]. These models are able to give global descriptions of the general tendency of the changes of nuclear masses, whereas some of the parameters fail to elucidate inner physical interactions in nuclei. The third class are purely microscopic models, such as Hartree-Fock-Bogoliubov (HFB) models [27–32] and relativistic mean-field (RMF) models [33–40]. These models exploit several effective physical interactions, and can provide global descriptions on nuclear masses, of which the accuracy can almost match that of microscopic-macroscopic models [41]. However, recent researches in nuclear structure, nuclear decay, and nuclear reactions require higher accuracy of the predictions of nuclear masses.

Recently, in addition to the development of the theoretical mass models, statistical methods have been proposed to improve the descriptions of nuclear properties. The neural networks got employed in addressing different nuclear physics issues including many-body problems [42–45], radii predictions [46–48], decay descriptions [49–52], reactions [53–56], and especially nuclear mass systematics [57–63]. There are a series of advantages of neural networks in predicting nuclear masses. First of all, compared with other traditional statistical methods, neural networks do not require a formal fitting function, which makes them more flexible and

*liujian@upc.edu.cn

accessible. Secondly, complex nonlinear mass relationships can be incorporated in a neural network with adjustable connection weights for predicting nuclear masses. Finally, a neural network belongs to the class of universal approximators, which can make global predictions for nuclear masses. Despite these advantages, there are two non-negligible disadvantages of neural networks. First, the neural network is a "black box", which makes it difficult to study any relationship deeply in the data [64]. Besides, the structures of neural networks are usually complicated, which requires a large amount of computations.

In order to overcome these disadvantages, some machine learning algorithms different from neural networks have also been applied to predict nuclear properties. In a recent paper, a naive Bayesian probabilities (NBP) classifier was introduced to refine the nuclear charge radii, which regarded predictions as a classification problem [65]. By constructing a naive Bayesian classifier, nuclear charge radii can be well described. For certain physical problems which can be converted to classification tasks, the naive Bayesian algorithm can perform better than other more sophisticated learning schemes [66]. It is the special zero-one loss function of the naive Bayesian method that is responsible for the surprising performances [66–68]. Moreover, due to the simple structure of the method, the naive Bayesian method is capable of uncovering the inner relationships in the data, which is helpful for the physical analysis. Additionally, the amount of computation for the naive Bayesian algorithm is much smaller than that for other machine learning schemes.

In this paper, an improved naive Bayesian probabilities (iNBP) classifier is put forward to describe the nuclear masses. Compared with the classical Bayesian formula, there are two innovations in the iNBP method. One is that a more appropriate classification is given by the *k-means* algorithm, and the other is that local relationships between nuclear masses are considered by introducing the weight functions. The key strategy of the iNBP method is to turn the prediction of nuclear masses to a classification problem. Based on certain nuclear mass models, the mass residuals $\delta(Z, N)$, which represent the deviations between experimental data and theoretical results, are provided initially for each nucleus. Then, the residuals $\delta(Z, N)$ are clustered into several groups by the *k-means* methods. For each group, there is a classification value δ_i defined as the mean value of residuals in the group. The original residuals $\delta(Z, N)$ in the group are recalibrated as the classification value δ_i. Next, we calculate the posterior probability of different classification values δ_i based on the Bayesian formula, and choose the value δ_i with the maximum probability as the estimated residual. During the calculations, the local relations between the neighboring nuclei are taken into account by introducing weight functions. Finally, the raw results of theoretical models are refined by adding the predicted residuals from the iNBP method.

To examine its effectiveness, the iNBP method is applied to three different types of sophisticated nuclear models: the microscopic-macroscopic FRDM model, the nonrelativistic HFB model, and the relativistic RMF model in this paper. Both the interpolating and extrapolating capabilities of the iNBP method are evaluated. The nuclei with the proton number $Z \geqslant 8$ are analyzed. For the interpolation, the data set includes 3245 nuclei between ^{16}O and ^{295}Og in the atomic mass evaluation of 2016 (AME2016) [69]. For the extrapolation, the learning set consists of 3007 nuclei in the atomic mass evaluation of 2003 (AME2003) [70], while the validation set includes 238 nuclei which got newly added in the AME2016. In addition, we display the corrections of the binding energies for Ca outside of the AME compilation by a graphic depiction. Our results illustrate that the iNBP method has good interpolating and extrapolating capabilities, and can be applied to predict nuclear masses in unknown regions of the nuclear chart.

This paper is organized as follows. In the next section, the theoretical framework of the iNBP classifier is discussed in detail. The results of predictions and corresponding discussions are presented in Sec. III. Finally, in Sec IV a brief summary is given.

II. FORMALISM

The iNBP method involves two steps. In the first step, we use the k-means algorithm to classify all the residuals in certain groups, and the residuals are recalibrated as the corresponding classification values. In the second step, the probabilities of all the classification values are calculated by the Bayesian formula for the predicted nucleus, and the estimated residual of the predicted nucleus is obtained from these probabilities. Assessment criteria of accuracy and uncertainties of iNBP method are also presented in this section.

A. Classification by the *k*-means algorithm

In this step, we choose the k-means algorithm to classify all the raw residuals. The k-means algorithm automatically divides a data set into k groups by an iterative refinement technique [71,72]. Starting from generating k initial group centers $\delta_i^{(0)}$ $(i = 1, ..., k)$ randomly, the Euclidean distances between each residual $\delta(Z, N)$ and each group center $\delta_i^{(0)}$ are calculated. Then, the residual is assigned to the group with the nearest group center. Next, the new group center is recalculated as the mean value of the residuals in this group:

$$\delta_i^{(t)} = \frac{1}{N_i} \sum_{c_i} \delta(Z, N)|_{c_i}, \tag{1}$$

where c_i represents the group label. $\delta(Z, N)|_{c_i}$ is the residual in the group c_i, and N_i indicates the number of the residuals in this group. By iterative calculations with the procedure mentioned above, the group centers can converge to final stable classification values δ_i, and the assignment of the residuals to groups is not further changed [71]. The residuals $\delta(Z, N)$ in each group are recalibrated as the final classification value δ_i.

In Table I, we present an example for the classification with the k-means method, in which the residuals of 3245 nuclei are separated into ten groups. The theoretical mass values are calculated by the RMF model with NL3* parameter set [73], and the experimental data are taken from the AME2016 [69]. The intervals of these ten groups are shown in Table I with their classification values δ_i displayed. Residuals

TABLE I. An example of the classification by the k-means algorithm with ten groups and the corresponding classification values δ_i. The number of the residuals in each group is given as well.

Intervals (MeV)	δ_i (MeV)	N_i
$(-\infty, -12.48)$	$\delta_1 = -15.8$	17
$(-12.48, -7.00)$	$\delta_2 = -8.91$	82
$(-7.00, -4.03)$	$\delta_3 = -5.01$	277
$(-4.03, -2.24)$	$\delta_4 = -3.05$	404
$(-2.24, -0.70)$	$\delta_5 = -1.42$	440
$(-0.70, 0.67)$	$\delta_6 = 0.03$	496
$(0.67, 1.91)$	$\delta_7 = 1.31$	558
$(1.91, 3.16)$	$\delta_8 = 2.51$	494
$(3.16, 4.74)$	$\delta_9 = 3.81$	358
$(4.74, +\infty)$	$\delta_{10} = 5.68$	119

$\delta(Z, N) = M_{\rm exp} - M_{\rm th}$ in each interval are recalibrated as the corresponding δ_i.

B. Prediction from the iNBP classifier

Based on the classification table presented in Table I, the residuals $\delta(Z_t, N_t)$ for certain nuclei with proton number Z_t and neutron number N_t are predicted by the iNBP classifier in this part. Supposing the proton numbers and neutron numbers are independent, the posterior probability $P(\delta_i|\, Z_t, N_t)$ of a certain classification value δ_i given Z_t and N_t can be calculated from the prior probabilities $P(Z_t)$, $P(N_t)$, $P(\delta_i)$ and the conditional probabilities $P(Z_t|\, \delta_i)$, $P(N_t|\, \delta_i)$ with the Bayesian formula [65]

$$P(\delta_i|\, Z_t, N_t) = \frac{P(Z_t, N_t|\, \delta_i)P(\delta_i)}{P(Z_t, N_t)} = \frac{P(Z_t|\, \delta_i)P(N_t|\, \delta_i)P(\delta_i)}{P(Z_t)P(N_t)}. \tag{2}$$

The prior probabilities $P(Z_t)$, $P(N_t)$, and $P(\delta_i)$ represent the occurrence frequencies of features Z_t, N_t, and δ_i in the sample, respectively. The conditional probabilities $P(Z_t|\, \delta_i)$ and $P(N_t|\, \delta_i)$ represent the occurrence frequencies of features Z_t and N_t in the group with the classification values δ_i.

In order to take into account the local relations between the neighboring nuclei, a weight function is further introduced in calculating the prior and conditional probabilities:

$$w(Z, N) = \exp\left[-\frac{(Z - Z_t)^2 + (N - N_t)^2}{2\rho^2}\right]. \tag{3}$$

The weight function Eq. (3) models the spatial dependence of the nearby nuclei. There is only one parameter ρ in Eq. (3). For different predicted nuclei, the parameter ρ is the same. In Eq. (4) of Ref. [63], an exponential quadratic covariance kernel is defined in Gaussian processes (GP) to take into account the correlations between the masses of different nuclei. In this paper, the role of the weight function of the iNBP method is similar to the kernel function of the GP method in Ref. [63].

During the studies, the variance ρ of Eq. (3) is set to be $\rho = 3$. Then the prior probabilities $P(Z_t)$, $P(N_t)$, $P(\delta_i)$ in the Bayesian formula Eq. (2) are converted to $P_{\rm wt}(Z_t)$, $P_{\rm wt}(N_t)$, and $P_{\rm wt}(\delta_i)$:

$$\begin{aligned}
P(Z_t) &= \frac{\sum_{Z,N} \delta_{Z,Z_t}}{\sum_{Z,N} 1} \longrightarrow P_{\rm wt}(Z_t) = \frac{\sum_{Z,N} \delta_{Z,Z_t} \times w(Z, N)}{\sum_{Z,N} w(Z, N)}, \\
P(N_t) &= \frac{\sum_{Z,N} \delta_{N,N_t}}{\sum_{Z,N} 1} \longrightarrow P_{\rm wt}(N_t) = \frac{\sum_{Z,N} \delta_{N,N_t} \times w(Z, N)}{\sum_{Z,N} w(Z, N)}, \\
P(\delta_i) &= \frac{\sum_{\delta=\delta_i} 1}{\sum_{Z,N} 1} \longrightarrow P_{\rm wt}(\delta_i) = \frac{\sum_{\delta=\delta_i} w(Z, N)}{\sum_{Z,N} w(Z, N)},
\end{aligned} \tag{4}$$

where δ_{Z,Z_t} is the Kronecker δ function. The summation $\sum_{Z,N} 1$ represents the amount of the nuclei in the sample set, and the summation $\sum_{\delta=\delta_i} 1$ represents the number of nuclei in the group with classification values $\delta = \delta_i$.

Similarly, the expressions of the conditional probabilities $P(Z_t|\, \delta_i)$, $P(N_t|\, \delta_i)$ are also converted to $P_{\rm wt}(Z_t|\, \delta_i)$, $P_{\rm wt}(N_t|\, \delta_i)$:

$$\begin{aligned}
P(Z_t|\, \delta_i) &= \frac{\sum_{\delta=\delta_i} \delta_{Z,Z_t}}{\sum_{\delta=\delta_i} 1} \longrightarrow P_{\rm wt}(Z_t|\, \delta_i) \\
&= \frac{\sum_{\delta=\delta_i} \delta_{Z,Z_t} \times w(Z, N)}{\sum_{\delta=\delta_i} w(Z, N)}, \\
P(N_t|\, \delta_i) &= \frac{\sum_{\delta=\delta_i} \delta_{N,N_t}}{\sum_{\delta=\delta_i} 1} \longrightarrow P_{\rm wt}(N_t|\, \delta_i) \\
&= \frac{\sum_{\delta=\delta_i} \delta_{N,N_t} \times w(Z, N)}{\sum_{\delta=\delta_i} w(Z, N)}.
\end{aligned} \tag{5}$$

The parameter ρ in Eq. (3) can describe the contributions of the adjacent nuclei to the target nucleus. The smaller the value of ρ, the greater the contributions of the adjacent nuclei. When $\rho \to \infty$, the weight function $w(Z, N) \to 1$ in Eqs. (4) and (5), and all the nuclei contribute equally to the predicted nucleus in the Bayesian formula Eq. (2). The smaller the value of ρ, the higher for the improvement of the accuracy of the iNBP method. However, for too small ρ, the weight function Eq. (3) for certain predicted nucleus is very small, and in this case, the prediction cannot be made. Because for very small ρ, the prior probabilities $P(Z)$ or $P(N)$ are very close to zero for certain nuclei in Bayesian formula Eq. (2), which causes the posterior probabilities for different classification values δ_i cannot be calculated. By comprehensivly considering the prediction accuracy and the number of the predicted nuclei, we finally choose $\rho = 3$ in the paper. Detailed discussion is presented in Sec. III B to illustrate the influences of values of ρ on the prediction accuracy and the number of the predicted nuclei.

It should be noted that the law of total probability is also satisfied after introducing the weight functions:

$$\begin{aligned}
P_{\rm wt}(Z_t(N_t)) &= \sum P_{\rm wt}(Z_t(N_t)|\, \delta_i) P_{\rm wt}(\delta_i), \\
P_{\rm wt}(\delta_i) &= \sum P_{\rm wt}(\delta_i|\, Z_t(N_t)) P_{\rm wt}(Z_t(N_t)),
\end{aligned} \tag{6}$$

which reflects the fundamental relations between conditional probabilities Eqs. (5) and the prior probabilities Eqs. (4). The law of total probability Eqs. (6) ensures the adjusted posterior probability $P_{\rm wt}(\delta_i|\, Z_t, N_t)$ still satisfies the Bayesian

TABLE II. The prediction of residual for ^{29}Si with $Z_t = 14$ and $N_t = 15$, based on ten classification values of Table I. The sample set contains 3245 nuclei with $Z \geqslant 8$, except the predicted nucleus ^{29}Si. The prior probabilities $P_{\rm wt}(\delta_i)$ and conditional probabilities $P_{\rm wt}(Z_t|\,\delta_i)$, $P_{\rm wt}(N_t|\,\delta_i)$ are also presented in this table, where the weight functions are introduced during the calculations.

δ_i (MeV)	$P_{\rm wt}(\delta_i)$	$P_{\rm wt}(Z_t\,|\,\delta_i)$	$P_{\rm wt}(N_t|\,\delta_i)$	$P_{\rm wt}(\delta_i|\,Z_t, N_t)$
−15.8	0	0	0	0
−8.91	8.3×10^{-3}	0	0	0
−5.01	3.0×10^{-2}	0	0	0
−3.05	8.0×10^{-2}	5.1×10^{-2}	9.7×10^{-2}	1.8×10^{-2}
−1.42	4.3×10^{-2}	6.6×10^{-2}	6.4×10^{-2}	8.4×10^{-3}
0.03	0.16	9.7×10^{-2}	7.8×10^{-2}	5.5×10^{-2}
1.31	0.13	6.5×10^{-2}	0.21	8.4×10^{-2}
2.51	0.24	2.1×10^{-2}	0.23	5.4×10^{-2}
3.81	0.19	0.29	8.4×10^{-2}	0.22
5.68	0.12	0.32	0	0

formula. Combining Eqs. (4) and (5), the posterior probabilities $P_{\rm wt}(\delta_i|\,Z_t, N_t)$ can be calculated by the Bayesian formula Eq. (2). By comparing the posterior probabilities $P_{\rm wt}(\delta_i|\,Z_t, N_t)$ of each δ_i, the residual δ_m with the maximum $P_{\rm wt}(\delta_m|\,Z_t, N_t)$ is regarded as the predicted residual of the nucleus (Z_t, N_t).

As an example of the iNBP method, we refine the theoretical nuclear mass of ^{29}Si to illustrate the prediction process. There are 3245 nuclei with $Z \geqslant 8$ in AME2016 whose experimental masses are known. The raw residuals $\delta(Z, N)$ are calculated by the RMF model with the NL3* parameter set [73]. The sample set contains 3245 nuclei in AME2016, except ^{29}Si itself. All the residuals $\delta(Z, N)$ are grouped into ten classes in Table I, and recalibrated as representing classification values δ_i. The prior probabilities $P_{\rm wt}(\delta_i)$, $P_{\rm wt}(Z_t = 14)$, $P_{\rm wt}(N_t = 15)$ and the conditional probabilities $P_{\rm wt}(Z_t = 14|\,\delta_i)$, $P_{\rm wt}(N_t = 15|\,\delta_i)$ are calculated using Eqs. (4) and (5), and presented in Table II. Thereafter the posterior probability $P_{\rm wt}(\delta_i|\,Z_t = 14, N_t = 15)$ of each classification value δ_i are updated by the Bayesian formula Eq. (2), which is also presented in Table II.

One can see in Table II that the classification value $\delta_9 = 3.81$ MeV has the maximum posterior probability $P_{\rm wt}(\delta_9|\,Z_t = 14, N_t = 15) = 0.22$, which is much larger than those of other classification values. Therefore, the estimated residuals $\delta_{\rm est}$ of ^{29}Si are chosen to be $\delta_9 = 3.81$ MeV. For ^{29}Si, the raw theoretical binding energy from the RMF model is $B_{\rm raw} = 241.06$ MeV. By applying the iNBP method, the refined theoretical binding energy is $B_{\rm refi} = B_{\rm raw} + \delta_{\rm est} = 244.87$ MeV. Compared with the experimental data $B_{\rm exp} = 245.01$ MeV, we obtain a 96% improvement of the accuracy for ^{29}Si. It should be mentioned that, though all the data in the AME2016 are used as an input, most of the nuclei have little contributions to the prior and conditional probabilities in Eq. (2), due to the introduction of the weight function. For the ^{29}Si, about 5% of AME2016 data contribute to the final predictions.

C. Assessment criteria for model accuracy and uncertainties

The assessment criteria for the predicted results and the associated uncertainties are presented below. The standard deviation $\sigma_{\rm rms}$ can quantify the quality of the global deviations between theoretical results and experimental data for a certain model, and it is defined with

$$\sigma_{\rm rms}^2 = \frac{1}{X}\sum_{i=1}^{X}\left(M_{\rm exp}^i - M_{\rm th}^i\right)^2, \tag{7}$$

where X represents the total number of the nuclei in the data set. By comparing the standard deviations before and after the application of the iNBP method, we can evaluate the power of nuclear mass prediction of this method.

As a statistical model always produces a nondeterministic inference, the assessment for its uncertainty is essential. For the iNBP method, the uncertainties are contributed from the variety of the classification tables. The predictions in Table II are based on the classification table given in Table I. If we choose other classification tables, different results will be obtained. In this paper we choose 51 different classification tables whose number of groups varies from 10 to 60. The predicted nuclear mass $\langle M\rangle$ is the mean value of all estimated M obtained from 51 classifications tables. The uncertainty of the result ΔM is

$$\Delta M = \sqrt{\langle M^2\rangle - \langle M\rangle^2}, \tag{8}$$

where $\langle M^2\rangle$ is the mean value of the square of the mass.

III. RESULTS

In this section, the theoretical nuclear masses are refined following the iNBP method, with the raw results calculated within the semiempirical model FRDM, the nonrelativistic HFB model, and the relativistic RMF model, respectively. 51 classification tables are used to generate predictions. The global optimizations are carried out and the extrapolating capabilities of the iNBP method are assessed. Moreover, to intuitively test the validity of the iNBP method, we present the corrections for the nuclear binding energies of the Ca isotopic chains outside of the AME compilation.

A. Global optimization of iNBP method

We first analyze the performances of the iNBP method in global optimizations. 3245 nuclei with $Z \geqslant 8$ in the AME2016 [69] are chosen as the entire set. The theoretical binding energies for each nuclei in the entire set are calculated by the FRDM model, the HFB model with UNEDF1 parameter set [74], and the RMF model with NL3* parameter set [73], respectively. Then the corresponding raw residual $\delta_{\rm raw}(Z, N)$ for each nucleus is obtained. In Table III, we present the standard deviations $\sigma_{\rm pre}$ of the theoretical masses for 3245 nuclei in the entire set.

After the calculations of the nuclear masses, the iNBP method is further applied to refine the theoretical results. The sample set for each target nucleus includes 3245 nuclei in AME2016 except the target nucleus itself. Firstly, each residual is classified into a certain group by the k-means

TABLE III. The standard deviation $\sigma_{\rm pre}$ (MeV) from theoretical models and $\sigma_{\rm post}$ (MeV) from the iNBP method. 3245 nuclei in AME2016 with $Z \geqslant 8$ are chosen as the entire set.

Models	FRDM	HFB	RMF
$\sigma_{\rm pre}$	0.89	2.11	3.35
$\sigma_{\rm post}$	0.33	0.56	1.34
$\Delta\sigma/\sigma_{\rm pre}$	63%	73%	60%

algorithm and recalibrated as a corresponding classification value δ_i. For the targeted nucleus (Z_t, N_t), the prior probabilities $P_{\rm wt}(Z_t(N_t))$, $P_{\rm wt}(\delta_i)$ and the conditional probabilities $P_{\rm wt}(Z_t(N_t)|\,\delta_i)$ are calculated with Eqs. (4) and (5). With these probabilities, the posterior probabilities $P_{\rm wt}(\delta_i|\,Z_t, N_t)$ are given for each classification values δ_i by the Bayesian formula Eq. (2). The classification value δ_m with the maximum posterior probability is identified as the estimated residual of the target nucleus. Repeating this procedure 51 times with different classification tables and averaging the estimated residuals, we can obtain the final predicted residual of the targeted nucleus. Adding the predicted residual to the raw theoretical nuclear mass, we finally obtain the corrected nuclear mass.

We also calculate the $\sigma_{\rm post}$ for the corrected nuclear masses from the iNBP method. The power of the global optimization of the iNBP method can be quantified as the relative change of the standard deviation, $\Delta\sigma/\sigma_{\rm pre} = (\sigma_{\rm pre} - \sigma_{\rm post})/\sigma_{\rm pre}$. In Table III one can see that convincing improvements are obtained with the iNBP refinement. For the FRDM model, the standard derivation is $\sigma_{\rm post} = 0.33$ MeV after the iNBP corrections, and the accuracy of the descriptions on the nuclear mass improves by 63%. For the HFB model, after the iNBP refinements the standard derivation $\sigma_{\rm post}$ becomes 0.56 MeV, and the accuracy of the descriptions on the nuclear mass improves by 73%. For the RMF model, the standard deviation is improved to $\sigma_{\rm post} = 1.34$ MeV, with a 60% reduction in σ after the iNBP refinements.

To illustrate graphically the performance of the iNBP method in refining mass predictions, in Fig. 1 we compare the raw residuals and the corrected residuals of 3245 nuclei in the entire set. Raw residuals from the FRDM, HFB, and RMF models are displayed on the left panels and the corrected residuals from the iNBP method are on the right panels. It can be seen from Fig. 1 that the iNBP method greatly improves the accuracy of the descriptions of the nuclear masses, especially on the region of heavy nuclei for the FRDM model, and the nuclei near the drip line for the HFB and RMF models.

Overall, by combining the iNBP method with the empirical and microscopic models and utilizing data, we arrive at a stronger predictive power for nuclear masses. This suggests that the iNBP method incorporates the mass information of nuclei with the same Z or N in a statistical way, and provides necessary and rational corrections to the results from mean-field theories. However, it should be noticed that there is a limitation of the iNBP method. Specifically, predictions cannot be made for nuclei with the proton number or neutron number absent from the sample set, because the prior probability $P(Z)$ or $P(N)$ is zero in these cases. There are 3245 nuclei in the entire set. For the FRDM model, masses of 3240 nuclei can be predicted. For the HFB and RMF models, the number of nuclei for which predictions can be made is 3242. For the nuclei unpredicted by the iNBP method, the prior probability $P_{\rm wt}(\delta_i)$ is regarded as the posterior probability $P_{\rm wt}(\delta_i|\,Z_t, N_t)$ as an approximation.

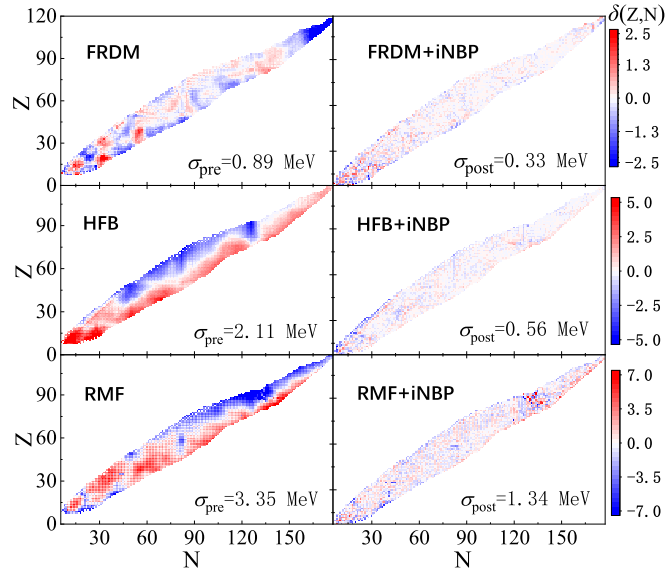

FIG. 1. Left panels: Raw residuals $\delta_{\rm raw}$ of theoretical nuclear masses for the FRDM model, the HFB model, and the RMF model with respect to the entire set, which includes 3245 nuclei in AME2016. Right panels: The corresponding corrected residuals $\delta_{\rm corr}$ of nuclear masses by the iNBP method. The values of the standard deviations before and after the iNBP refinements on the entire set are also presented in the figure.

B. Extrapolating capabilities of the iNBP method

It is essential to investigate the capabilities of extrapolation of the iNBP method, since extrapolation is more challenging than interpolation. In this part, we analyze the extrapolating capabilities of the iNBP method and illustrate the effect of the value of ρ. Before the extrapolation, all 3245 nuclei with $Z \geqslant 8$ in the entire set were assigned to the learning set and the validation set. The learning set includes 3007 nuclei in the AME2003 [70], and the validation set consists of the newly added 238 nuclei in the AME2016 [69].

First, we assess the extrapolating capability of the iNBP method based on the FRDM model. In Table IV we provide the raw standard deviations $\sigma_{\rm pre}$ calculated by the FRDM model for the learning set and the validation set. For the learning set the deviation $\sigma_{\rm pre}$ is 0.87 MeV, and for the validation set the deviation $\sigma_{\rm pre}$ is 1.16 MeV. This indicates that the raw FRDM model extrapolates well. Then we employ the iNBP method to make predictions for the validation set, where the prior and the conditional probabilities are obtained from data in the learning set. In Table IV we display the standard deviations $\sigma_{\rm post}$ for the learning set and the validation set after the iNBP corrections. The relative change of the standard

TABLE IV. The standard deviation $\sigma_{\rm pre}$ (MeV) of nuclear masses obtained from the raw FRDM, HFB, and RMF models, and the standard deviation $\sigma_{\rm post}$ (MeV) after the iNBP corrections. The entire set includes 3245 nuclei with $Z \geqslant 8$ in the AME2016 compilation. The learning set includes 3007 nuclei with $Z \geqslant 8$ in AME2003 compilation, and the validation set includes newly added 238 nuclei in AME2016 compilation.

	Learning set			Validation set			Entire set		
Models	FRDM	HFB	RMF	FRDM	HFB	RMF	FRDM	HFB	RMF
$\sigma_{\rm pre}$	0.87	2.08	3.20	1.16	2.45	4.90	0.89	2.11	3.35
$\sigma_{\rm post}$	0.34	0.56	1.38	0.50	1.06	2.81	0.33	0.56	1.34
$\Delta\sigma/\sigma_{\rm pre}$	61%	73%	57%	57%	57%	43%	63%	73%	60%

deviation $\Delta\sigma/\sigma_{\rm pre}$ is used to quantify the improvements of the accuracy of the mass predictions. From Table IV it can be seen that by applying the iNBP method on the FRDM model, the accuracy of the prediction for the validation set improves by 57%. Compared with the improvement for the learning set (61% for the standard deviation reduction), one can see the iNBP method has a reliable extrapolating capability.

Besides the FRDM model, the extrapolating capabilities of the iNBP method are also discussed based on the HFB model and the RMF model. According to the calculations by the HFB model, for the learning set the raw standard deviation $\sigma_{\rm pre} = 2.08$ MeV, and for the validation set $\sigma_{\rm pre} = 2.45$ MeV, which illustrates the HFB model has the extrapolating ability. Similarly, the nuclear masses in the validation set are predicted by the iNBP method based on the learning set. For the validation set, we obtain a 57% improvement on the mass descriptions after the iNBP corrections from the relative change between $\sigma_{\rm pre}$ and $\sigma_{\rm post}$. Compared with the relative changes of the standard deviation of the learning set (73% improvement), we demonstrate a robust extrapolating ability for the iNBP method when applied to the HFB model.

For the RMF model, the raw prediction results are also presented in Table IV. For the learning set the raw standard deviation is $\sigma_{\rm pre} = 3.20$ MeV, and for the validation set it is $\sigma_{\rm pre} = 4.90$ MeV. This indicates the RMF model also has the extrapolating capability. With the iNBP refinements, we obtain a 57% reduction in the standard deviation for the learning set and a 43% reduction for the validation set. Compared with the relative changes between $\sigma_{\rm pre}$ and $\sigma_{\rm post}$ for the learning set and the validation set, one can see the iNBP method also has the definite extrapolating capability when applied on the RMF model.

Furthermore, we display the corrected results for the nuclei with relatively large raw residuals in Table V. For nuclei with the raw residuals $|\delta_{\rm raw}| > 12$ MeV in the validation set, we present the associated corrected residuals $|\delta_{\rm corr}|$. In Table V one can see impressive improvements in the mass predictions after the iNBP corrections. For example, the raw residual for ^{220}Np is $|\delta_{\rm raw}| = 12.60$ MeV. After the iNBP corrections, the corrected residual is $|\delta_{\rm corr}| = 3.22$ MeV, where a 74% improvement is obtained. This is because the inner mass relations between the neighboring nuclei are implicit in the iNBP method, unlike in the mean-field theories.

TABLE V. Nuclei in the validation set with raw residuals $|\delta_{\rm raw}| > 12$ MeV, calculated by the RMF model with NL3* parameter set, and the corrected residuals $|\delta_{\rm corr}|$, following the corrections from iNBP method. All residuals are in MeV.

Z	N	$\lvert\delta_{\rm raw}\rvert$	$\lvert\delta_{\rm corr}\rvert$
89	131	13.09	4.18
90	131	13.37	4.46
91	127	12.09	3.18
91	130	16.25	0.43
91	131	15.65	0.17
91	132	13.53	4.62
92	130	16.29	0.47
92	131	14.79	1.03
93	126	12.48	3.35
93	127	12.60	3.22
93	132	14.52	1.31
93	133	14.08	5.17

The results of Tables IV and V indicate that the iNBP method has good extrapolating capabilities. The improvements on the validation set show the stability of the iNBP method in extrapolations, and illustrate the rationality of the approach that incorporates local mass relations through the Bayesian formula. This improves the success of the iNBP method in predicting the masses of the unknown nuclei. There are also limitations for the iNBP extrapolations. As we already mentioned, if the proton number or the neutron number of a nucleus is absent from the learning set, the nuclear mass cannot be predicted by the iNBP method. There are 238 nuclei in the validation set. For the FRDM model, 229 corrected nuclear masses are obtained. For the HFB model, we obtain 234 corrected nuclear masses, and for the RMF model we obtain 233 corrected nuclear masses. For the unpredictable nuclei, we also use the prior probability $P_{\rm wt}(\delta_i)$ as the posterior probability $P_{\rm wt}(\delta_i|\, Z_t, N_t)$ as an approximation.

The value of ρ in weigh function Eq. (3) also influences the predicted abilities of the iNBP method. In Table VI we present the predicted abilities of different ρ^2 for the extrapolation in the case of the NL3* parameter set. One can see that as the value of ρ^2 decreases, the predicted accuracy of iNBP method is gradually improved. However, for too small ρ, the weight function can be very small, and in this case the prediction

TABLE VI. The predicted abilities of different ρ^2 in weight function for the extrapolation. The raw nuclear masses are calculated by the RMF model with NL3* parameter set. The learning set includes 3007 nuclei with $Z \geqslant 8$ in the AME2003 compilation, and the validation set includes newly added 238 nuclei in AME2016 compilation. $N_{\rm unpre}$ represents the number of nuclei which cannot be predicted by the iNBP method.

ρ^2	$\sigma_{\rm pre}$	$\sigma_{\rm post}$	$\Delta\sigma/\sigma_{\rm pre}$	$N_{\rm unpre}$
1	4.90	2.63	46%	16
3	4.90	2.67	45%	8
9	4.90	2.81	43%	5

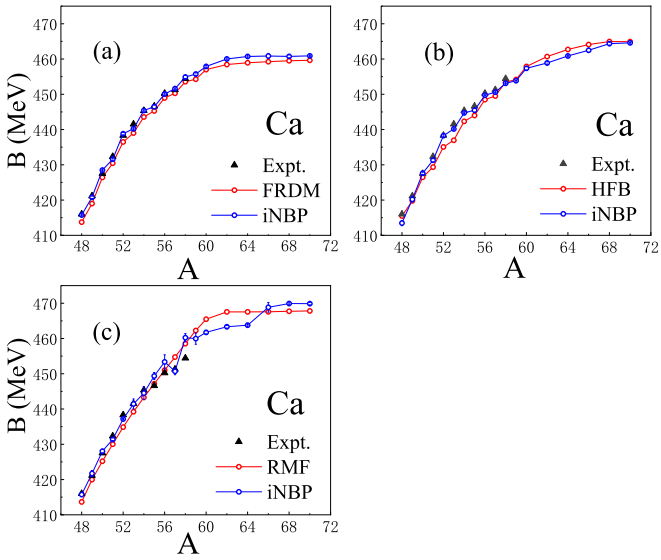

FIG. 2. (a) The theoretical and experimental binding energies for the Ca isotopes. The red circles represent raw results from the FRDM model and the blue circles represent the corrected results from the iNBP method. (b) The same as (a), but for the corrections on the results of HFB model. (c) The same as (a), but for the corrections on the results of RMF model.

cannot be made for certain predicted nucleus, which has been discussed in Sec. II B of this paper. By comprehensive considering the prediction accuracy and the number of the predicted nuclei for different nuclear models, we finally choose $\rho = 3$ in the paper.

C. Predictions of nuclear mass outside the AME compilation

As a follow-up to the previous discussions, we present a further illustration of the iNBP refinements by considering nuclear chains outside the AME compilation. The neutron-rich Ca isotopes are chosen as candidates. At present, the heaviest Ca isotopes observed is ^{60}Ca [75], and the predicted drip line is up to ^{70}Ca [62]. Compared with the odd-A nuclei, the even-even nuclei are more bound, and the $2n$ drip line extends further away. With the 3245 nuclei of $Z \geqslant 8$ in AME2016 as the learning set, the binding energies of even-even isotopes of $^{58-70}$Ca are predicted and presented in Fig. 2. The raw theoretical values obtained by the FRDM model [23], the HFB model with UNEDF1 parameter set [74], and the RMF model with NL3* parameter set [73] are also presented in the figures.

In Fig. 2(a), the raw binding energies of the isotopes $^{47-57}$Ca of the FRDM model are smaller than the experimental data. However, after the iNBP refinements, the corrected binding energies can well reproduce the experimental data, which can be attributed to the local mass correlations brought about from the Bayesian formula Eq. (2). From Figs. 2(b) and 2(c), one can also see that satisfactory improvements are obtained after the iNBP corrections on the raw results of HFB and RMF models. This also shows the considerable effects of the embedded local mass relations on the descriptions of nuclear binding energies.

In order to exhibit the validity of iNBP method, the extrapolations towards the neutron drip line are carried out for the FRDM, HFB, and RMF models in Figs. 2. From the calculations of three base models, the raw binding energies of ^{68,70}Ca are almost equal to each other, which means the $2n$ drip line located around the ^{70}Ca in theoretical predictions. However, the binding energies of $^{60-70}$Ca obtained from the HFB model are nearly 5 MeV smaller than those from the FRDM and RMF models. After the iNBP refinements, the predicted nuclear mass of isotopes $^{60-64}$Ca for the three models are very close to each other, which means the extrapolations outside the AME compilation with the different base models are in agreement. For ^{68,70}Ca, its neutron number is too far from the AME data set, and the impacts of the raw results are much larger than existing experimental data in AME compilation. This causes the corrected results of RMF model is still larger than those of FRDM and HFB models.

The essence of the iNBP method can be reflected in Fig. 2. The sophisticated semiempirical models and mean-field theories convincingly describe the global changing trends of nuclear binding energies, and the descriptions can be systematically fine-tuned by the iNBP method. The extrapolations in Fig. 2 further suggest that the iNBP method captures effects of residual interactions missing in the FRDM and mean-field approaches, which results in visible corrections on the binding energies. Therefore, the iNBP method can be applied to predict the nuclear mass outside the AME tables. In follow-up studies, we will further take into account more results of different energy density functionals and consider the impacts of model mixing to study the extent of bound nuclei.

IV. SUMMARY

Nuclear mass is a basic and important characteristic in nuclear physics. Many successful theoretical models have been proposed to describe nuclear masses, including the semiempirical models and the microscopic theories. Besides these, in recent years statistical methods such as neural networks have also been implemented to improve the descriptions on nuclear masses. In our previous work, we applied the naive Bayesian probability (NBP) classifier to refine the nuclear charge radii. In the present paper, an improved naive Bayesian probability (iNBP) classifier is further proposed to address the predictions of nuclear masses. There are two main innovations in the iNBP method. On one hand, a classification is given in terms of the k-means algorithm; on the other hand, the local mass relationships between neighboring nuclei are induced with the weight functions.

Global optimizations and extrapolating capabilities have been discussed in assessing the effectiveness of the iNBP method. First, we analyze the global optimizations of the method. Our results demonstrate that the iNBP method can offer considerable and reasonable corrections within the global description of nuclear masses. Second, the extrapolating capabilities of the iNBP method are analyzed. We use the iNBP method to predict the nuclear masses of the nuclei newly added in AME2016, based on the data in AME2003. Compared the relative changes of the standard deviation $\Delta\sigma/\sigma_{\rm pre}$

for the learning set and the validation set, one can see the iNBP method has reliable extrapolating capabilities. Finally, the iNBP corrections for the isotopes outside the AME compilation are exhibited in Fig. 2 to provide qualitative insights. The extrapolations with the different base models outside the AME compilation are in agreement with each other, which indicates the iNBP method can predict nuclear masses in the unknown regions. Combining the global optimizations and extrapolating analysis, we find that the iNBP method can provide necessary fine adjustments on the robust trends of binding energies from sophisticated nuclear models, and the corrected binding energies well reproduce the experimental data.

With the application of the iNBP method, the description of the nuclear masses is significantly improved. This is due to the local relations statistically taken into account by the iNBP method. Theoretical nuclear models convincingly describe the principal changing trends of nuclear masses, and iNBP method provides sizable and reliable fine-tuning. In iNBP method, the Bayesian formula reflects the statistical mass relations of the nuclei with the same Z and N, and the weight functions include the local mass correlations between the neighboring nuclei. The extrapolation results obtained from the iNBP method can provide support for the experimental studies on masses of exotic nuclei near the drip line. The iNBP method proposed in this paper can also be used to study other nuclear properties such as nuclear decay, nuclear charge radii, and nuclear reactions.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 11505292, 11605105, 11822503, 11975167, and 12035011), by the Shandong Provincial Natural Science Foundation, China (Grant No. ZR2020MA096), by the Fundamental Research Funds for the Central Universities (Grant No. 20CX05013A), and by the Graduate Innovative Research Funds of China University of Petroleum (East China) (Grant No. YCX2020104), and by the U.S. Department of Energy Office of Science under Grant No. DE-SC0019209.

[1] D. Lunney, J. M. Pearson, and C. Thibault, Rev. Mod. Phys. **75**, 1021 (2003).
[2] S. Ćwiok, P.-H. Heenen, and W. Nazarewicz, Nature (London) **433**, 705 (2005).
[3] B. Sun, R. Knöbel, Y. A. Litvinov, H. Geissel, J. Meng *et al.*, Nucl. Phys. A **812**, 1 (2008).
[4] S. Péru and M. Martini, Eur. Phys. J. A **50**, 88 (2014).
[5] D. Bai and Z. Ren, Phys. Rev. C **101**, 034311 (2020).
[6] J. Giovinazzo, B. Blank, M. Chartier, S. Czajkowski, A. Fleury, M. J. Lopez Jimenez, M. S. Pravikoff, J.-C. Thomas, F. de Oliveira Santos, M. Lewitowicz, V. Maslov, M. Stanoiu, R. Grzywacz, M. Pfützner, C. Borcea, and B. A. Brown, Phys. Rev. Lett. **89**, 102501 (2002).
[7] M. Pfützner, E. Badura, C. Bingham, B. Blank, M. Chartier *et al.*, Eur. Phys. J. A **14**, 279 (2002).
[8] N. Wan, C. Xu, and X. Zhang, J. Phys. G **47**, 055107 (2020).
[9] P. Möller, D. G. Madland, A. J. Sierk, and A. Iwamoto, Nature (London) **409**, 785 (2001).
[10] S. Panebianco, J.-L. Sida, H. Goutte, J.-F. Lemaître, N. Dubray, and S. Hilaire, Phys. Rev. C **86**, 064601 (2012).
[11] D. Regnier, N. Dubray, N. Schunck, and M. Verrière, Phys. Rev. C **93**, 054611 (2016).
[12] S. Goriely, N. Chamel, and J. M. Pearson, Phys. Rev. C **82**, 035804 (2010).
[13] D. Martin, A. Arcones, W. Nazarewicz, and E. Olsen, Phys. Rev. Lett. **116**, 121101 (2016).
[14] P.-C. Chu, Y. Zhou, X.-H. Li, and Z. Zhang, Phys. Rev. D **100**, 103012 (2019).
[15] J. Hakala, J. Dobaczewski, D. Gorelov, T. Eronen, A. Jokinen, A. Kankainen, V. S. Kolhinen, M. Kortelainen, I. D. Moore, H. Penttilä, S. Rinta-Antila, J. Rissanen, A. Saastamoinen, V. Sonnenschein, and J. Äystö, Phys. Rev. Lett. **109**, 032501 (2012).
[16] B. Franzke, H. Geissel, and G. Münzenberg, Mass Spectrom. Rev. **27**, 428 (2008).
[17] R. Kalpakchieva, H. G. Bohlen, W. von Oertzen, B. Gebauer, M. von Lucke-Petsch, T. N. Massey, A. N. Ostrowski, T. Stolla, M. Wilpert, and T. Wilpert, Eur. Phys. J. A **7**, 451 (2000).
[18] B. H. Sun, Y. A. Litvinov, I. Tanihata, and Y. H. Zhang, Front. Phys. **10**, 1 (2015).
[19] G. T. Garvey and I. Kelson, Phys. Rev. Lett. **16**, 197 (1966).
[20] J. Barea, A. Frank, J. G. Hirsch, P. Van Isacker, S. Pittel, and V. Velázquez, Phys. Rev. C **77**, 041304(R) (2008).
[21] J. Piekarewicz, M. Centelles, X. Roca-Maza, and X. Viñas, Eur. Phys. J. A **46**, 379 (2010).
[22] Y. Y. Cheng, Y. M. Zhao, and A. Arima, Phys. Rev. C **89**, 061304(R) (2014).
[23] P. Möller, A. J. Sierk, T. Ichikawa, and H. Sagawa, At. Data Nucl. Data Tables **109-110**, 1 (2016).
[24] N. Wang, M. Liu, and X. Wu, Phys. Rev. C **81**, 044322 (2010).
[25] M. Liu, N. Wang, Y. Deng, and X. Wu, Phys. Rev. C **84**, 014333 (2011).
[26] Y. Y. Zong, C. Ma, Y. M. Zhao, and A. Arima, Phys. Rev. C **102**, 024302 (2020).
[27] L. Wang, J. Liu, T. Liang, Z. Ren, C. Xu, and S. Wang, J. Phys. G **47**, 025105 (2020).
[28] E. Perlińska, S. G. Rohoziński, J. Dobaczewski, and W. Nazarewicz, Phys. Rev. C **69**, 014316 (2004).
[29] N. Schunck, J. Dobaczewski, J. McDonnell, J. Moré, W. Nazarewicz, J. Sarich, and M. V. Stoitsov, Phys. Rev. C **81**, 024316 (2010).
[30] J. C. Pei, A. T. Kruppa, and W. Nazarewicz, Phys. Rev. C **84**, 024311 (2011).
[31] S. Goriely, N. Chamel, and J. M. Pearson, Phys. Rev. C **88**, 024308 (2013).
[32] J. Dobaczewski, A. V. Afanasjev, M. Bender, L. M. Robledo, and Y. Shi, Nucl. Phys. A **944**, 388 (2015).
[33] L. Wang, J. Liu, R. Wang, M. Lyu, C. Xu, and Z. Ren, Phys. Rev. C **103**, 054307 (2021).
[34] J. Liu, C. Xu, S. Wang, and Z. Ren, Phys. Rev. C **96**, 034314 (2017).

[35] P. Ring, Prog. Part. Nucl. Phys. **37**, 193 (1996).
[36] J. Meng, H. Toki, S. G. Zhou, S. Q. Zhang, W. H. Long, and L. S. Geng, Prog. Part. Nucl. Phys. **57**, 470 (2006).
[37] W.-C. Chen and J. Piekarewicz, Phys. Rev. C **90**, 044305 (2014).
[38] X. Roca-Maza, N. Paar, and G. Colò, J. Phys. G **42**, 034033 (2015).
[39] J. Liu, X. Zhang, C. Xu, and Z. Ren, Nucl. Phys. A **948**, 46 (2016).
[40] J. Liu, R. Xu, J. Zhang, C. Xu, and Z. Ren, J. Phys. G **46**, 055105 (2019).
[41] M. Bender, P.-H. Heenen, and P.-G. Reinhard, Rev. Mod. Phys. **75**, 121 (2003).
[42] G. Carleo and M. Troyer, Science **355**, 602 (2017).
[43] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nat. Phys. **14**, 447 (2018).
[44] D. Luo and B. K. Clark, Phys. Rev. Lett. **122**, 226401 (2019).
[45] S. Yoshida, Phys. Rev. C **102**, 024305 (2020).
[46] S. Akkoyun, T. Bayram, S. O. Kara, and A. Sinan, J. Phys. G **40**, 055106 (2013).
[47] R. Utama, W.-C. Chen, and J. Piekarewicz, J. Phys. G **43**, 114002 (2016).
[48] D. Wu, C. L. Bai, H. Sagawa, and H. Q. Zhang, Phys. Rev. C **102**, 054323 (2020).
[49] N. J. Costiris, E. Mavrommatis, K. A. Gernoth, and J. W. Clark, Phys. Rev. C **80**, 044332 (2009).
[50] Z. M. Niu, H. Z. Liang, B. H. Sun, W. H. Long, and Y. F. Niu, Phys. Rev. C **99**, 064307 (2019).
[51] U. B. Rodríguez, C. Z. Vargas, M. Gonçalves, S. B. Duarte, and F. Guzmán, J. Phys. G **46**, 115109 (2019).
[52] P. S. A. Freitas and J. W. Clark, arXiv:1910.12345.
[53] G. B. King, A. E. Lovell, L. Neufcourt, and F. M. Nunes, Phys. Rev. Lett. **122**, 232502 (2019).
[54] Z.-A. Wang, J. Pei, Y. Liu, and Y. Qiang, Phys. Rev. Lett. **123**, 122501 (2019).
[55] S. Akkoyun, Nucl. Instrum. Methods Phys. Res. B **462**, 51 (2020).
[56] C.-W. Ma, D. Peng, H.-L. Wei, Z.-M. Niu, Y.-T. Wang, and R. Wada, Chin. Phys. C **44**, 014104 (2020).
[57] S. Athanassopoulos, E. Mavrommatis, K. A. Gernoth, and J. W. Clark, Nucl. Phys. A **743**, 222 (2004).
[58] R. Utama, J. Piekarewicz, and H. B. Prosper, Phys. Rev. C **93**, 014311 (2016).
[59] R. Utama and J. Piekarewicz, Phys. Rev. C **96**, 044308 (2017).
[60] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, Phys. Rev. C **98**, 034318 (2018).
[61] R. Utama and J. Piekarewicz, Phys. Rev. C **97**, 014306 (2018).
[62] L. Neufcourt, Y. Cao, W. Nazarewicz, E. Olsen, and F. Viens, Phys. Rev. Lett. **122**, 062502 (2019).
[63] L. Neufcourt, Y. Cao, S. Giuliani, W. Nazarewicz, E. Olsen, and O. B. Tarasov, Phys. Rev. C **101**, 014319 (2020).
[64] J. V. Tu, J. Clin. Epidemiol. **49**, 1225 (1996).
[65] Y. Ma, C. Su, J. Liu, Z. Ren, C. Xu, and Y. Gao, Phys. Rev. C **101**, 014304 (2020).
[66] E. Frank, L. Trigg, G. Holmes, and I. H. Witten, Mach. Learn. **41**, 5 (2000).
[67] R. Kohavi and D. H. Wolpert, in *Proceedings of the Thirteenth International Conference on Machine Learning* (Santa Fe, NM, 1996), pp. 275–283.
[68] P. Domingos and M. Pazzani, Mach. Learn. **29**, 103 (1997).
[69] M. Wang, G. Audi, F. G. Kondev, W. J. Huang, S. Naimi, and X. Xu, Chin. Phys. C **41**, 030003 (2017).
[70] G. Audi, A. H. Wapstra, and C. Thibault, Nucl. Phys. A **729**, 337 (2003).
[71] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl, in *Proceedings of the Eighteenth International Conference on Machine Learning*, Vol. 1 (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001), pp. 577–584.
[72] A. K. Jain, Pattern Recogn. Lett. **31**, 651 (2010).
[73] G. A. Lalazissis, S. Karatzikos, R. Fossion, D. P. Arteaga, A. V. Afanasjev, and P. Ring, Phys. Lett. B **671**, 36 (2009).
[74] M. Kortelainen, J. McDonnell, W. Nazarewicz, P.-G. Reinhard, J. Sarich, N. Schunck, M. V. Stoitsov, and S. M. Wild, Phys. Rev. C **85**, 024304 (2012).
[75] O. B. Tarasov, D. J. Morrissey, A. M. Amthor, T. Baumann, D. Bazin, A. Gade, T. N. Ginter, M. Hausmann, N. Inabe, T. Kubo, A. Nettleton, J. Pereira, M. Portillo, B. M. Sherrill, A. Stolz, and M. Thoennessen, Phys. Rev. Lett. **102**, 142501 (2009).