Kolmogorov-Arnold networks in nuclear binding energy prediction

Hao Liu ⬤, Jin Lei ⬤,[*] and Zhongzhou Ren

School of Physics Science and Engineering, Tongji University, Shanghai 200092, China

This study explores the application of Kolmogorov-Arnold networks (KANs) in predicting nuclear binding energies, leveraging their ability to decompose complex multiparameter systems into simpler univariate functions. By utilizing data from the Atomic Mass Evaluation (AME2020) and incorporating features such as atomic number, neutron number, and shell effects, KANs achieved a significant lower root mean square error (0.26 MeV), surpassing traditional models. The symbolic regression analysis yielded simplified analytical expressions for binding energies, aligning with classical models like the liquid drop model and the Bethe-Weizsäcker formula. These results highlight KANs' potential in enhancing the interpretability and understanding of nuclear phenomena, paving the way for future applications in nuclear physics and beyond.

I. INTRODUCTION

The atomic nucleus, a quintessential quantum many-body system, exhibits remarkable structural complexity [1]. Binding energies (BE), synonymous with nuclear mass, are key characteristics of atomic nuclei and play an indispensable role in elucidating various nuclear phenomena. These phenomena include nuclear shapes, shell effects, pairing effects, and the emergence and disappearance of magic numbers [2–5]. Furthermore, binding energies are crucial in the synthesis of superheavy elements and in understanding nuclear astrophysical processes like the r-process [6] and x-ray bursts [7]. Therefore, both theoretical predictions and experimental measurements of nuclear binding energies are essential for advancements in nuclear physics.

Over the years, numerous theories and methods have been developed for predicting nuclear binding energies [1,2,8–21]. These include the Bethe-Weizsäcker formula [2,8], the Thomas-Fermi model [9], the Hartree-Fock-Bogoliubov mean field model [10], and *ab initio* methods [1,11,12]. These physics-based models have achieved commendable results within their respective applicable ranges. Concurrently, data-driven methods such as multilayer perceptrons [13–18], Gaussian processes [19,20], and support vector machines [20] have also found successful applications in the prediction of atomic nucleus masses.

However, traditional physical models necessitate a profound understanding of the inherent mechanisms of nuclear physical systems and face challenges when managing complex relationships. On the other hand, while data-driven machine learning methods can handle complex non-linear relationships, they often require substantial data and computational resources. In the data-rich field of nuclear physics, identifying intricate relationships among variables is a formidable

challenge. This predicament underscores the need for innovative approaches to handle the complexity of nuclear data and to extract meaningful insights.

Kolmogorov-Arnold networks (KANs) offer a promising solution [22]. KANs provide a simplified yet robust approach by decomposing complex multi-parameter systems into manageable univariate functions. This approach is based on the Kolmogorov-Arnold representation theorem [23,24], which asserts that any multivariate continuous function can be represented as a superposition of continuous univariate functions over a compact domain. This theoretical foundation allows KANs to handle complex, multi-parameter systems by breaking them down into simpler, univariate components.

One of the most compelling features of KANs in this scientific context is their capability for symbolic regression. Symbolic regression is a form of regression analysis that seeks mathematical expressions, in symbolic form, that best fit a given dataset. Unlike traditional numeric prediction models that offer a fixed structure and limited insights into underlying patterns, symbolic regression with KANs provides formulaic expressions that reveal the inherent relationships and governing laws within the data. This significantly enhances our understanding and offers a more profound and interpretable model.

The ability to derive symbolic equations directly from data can guide subsequent scientific research by offering clearer insights into the relationships and mechanics at play in nuclear phenomena. Such insights are invaluable for advancing theories, refining experimental designs, and developing new technologies in nuclear science. Hence, KANs do not merely predict but also illuminate the pathway towards a deeper comprehension of the intricate dynamics that define nuclear structures.

In the broader context of artificial intelligence (AI) applications in nuclear physics, there is a growing interest in leveraging machine learning to study the intricate phenomena of nuclear physics and derive formulas directly from data. The

---

*Contact author: jinl@tongji.edu.cn

024316-1

liquid drop model, which provides a macroscopic description of the nucleus, has been instrumental in understanding nuclear binding energies. AI techniques, particularly those involving deep learning and symbolic regression, have the potential to learn from vast amounts of nuclear data and derive similar or even more refined mass formulas. These data-driven approaches can capture subtle patterns and correlations that might be overlooked by traditional models, thus offering a complementary perspective to theoretical frameworks. A significant advantage of AI for physics is its ability to uncover hidden relationships within complex datasets. In the current study, we aim to derive mass formulas using AI, harnessing its power to reveal insights that can enhance our understanding of nuclear binding energies.

In this study, we aim to explore the potential of KANs in predicting nuclear binding energies. By leveraging the capabilities of KANs, we seek to bridge the gap between data-driven insights and theoretical understanding, ultimately contributing to the advancement of nuclear science. The paper is organized as follows: In Sec. II, we provide a brief overview of the Kolmogorov-Arnold network and the data collection and feature selection used in the current study. Sec. III presents the application of KANs to nuclear binding energy. Finally, we conclude with a summary and discussion in Sec. IV.

II. FORMALISM

A. Kolmogorov-Arnold network

The Kolmogorov-Arnold theorem asserts that any continuous multivariate function on a bounded domain can be decomposed into a finite composition of continuous univariate functions and addition. Specifically, for a continuous function $f : [0, 1]^n \rightarrow \mathbb{R}$:

$$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right). \quad (1)$$

Here, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is an n-dimensional vector with each x_i in [0, 1]. The functions $\phi_{q,p} : [0, 1] \rightarrow \mathbb{R}$ and $\Phi_q : \mathbb{R} \rightarrow \mathbb{R}$ are continuous univariate functions. This decomposition shows that addition is the only truly multivariate operation, as all other multivariate functions can be built using univariate functions and sums. This is useful for identifying learnable and explainable patterns in data. However, practical challenges in machine learning arise due to the potential non-smooth and fractal nature of the functions $\phi_{q,p}$ and Φ_q, complicating decomposition and learning processes.

Recent research by Liu *et al.* [22] has renewed interest in the Kolmogorov-Arnold theorem and its use in neural networks. This has led to the creation of KANs. Liu *et al.* have improved the original framework by allowing networks to have any number of layers and widths. This overcomes the original theorem's limitations, which restricted it to two layers of nonlinearities and a fixed number of terms ($2n + 1$). Since most real-world functions are usually smooth and have simple structures, these improvements help create more practical and efficient Kolmogorov-Arnold representations.
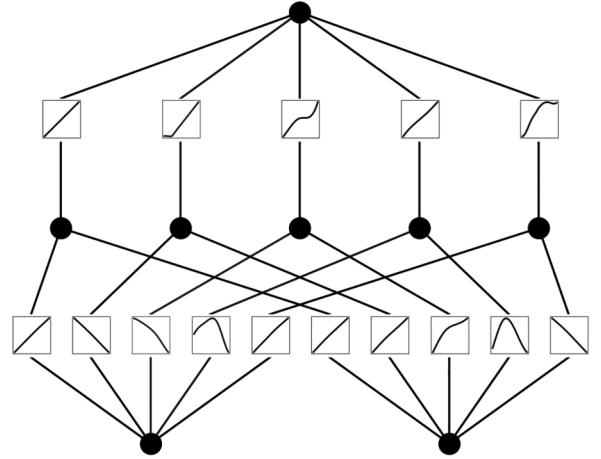

FIG. 1. Architecture of a shallow KAN, featuring two input features, a hidden layer with five nodes, and a single-node output layer.

In Fig. 1, a shallow KAN is illustrated. It consists of two input variables, x_1 and x_2, a hidden layer with five nodes, and a single output node. Each input is connected to the hidden nodes through edges characterized by adaptable functions, which replace the traditional fixed activation functions used in multilayer perceptrons (MLPs).

These adaptable functions are expanded into basis functions, specifically B-spline functions, which dynamically adjust based on the input data. The outputs from the hidden nodes are processed through additional spline-based edges, culminating in a summation at the output node. This structure allows for increasing the network depth, enabling the analysis of more complex physical phenomena.

B. Data collection and feature selection

In this study, we used mass excess values from the Atomic Mass Evaluation (AME2020) [25]. We focused on nuclei with both the atomic number (Z) and neutron number (N) greater than or equal to 8, covering 3456 nuclei. This selection ensures the model's precision[1] and stability for larger, more complex nuclei.

The experimental data are randomly divided into two subsets: 2,856 nuclei for training and 600 nuclei for testing. Figure 2 illustrates the selection of the training and test sets. The nuclei in the training and test sets remain consistent across all calculations.

To make the best use of our training set, we applied *k*-fold cross-validation [27–29], which ensures that every data point is used for both training and validation, maximizing the use of available data. Initially, we employed the Fisher-Yates shuffle algorithm [30] to ensure a thorough and unbiased randomization of the dataset, then split the dataset into k parts (folds).

---

[1]In our study, *accuracy* refers to a qualitative performance characteristic, expressing the closeness of agreement between a measurement result and the value of the measurand, whereas *precision* refers to the closeness of agreement between independent test results obtained under stipulated conditions, as discussed in Ref. [26].
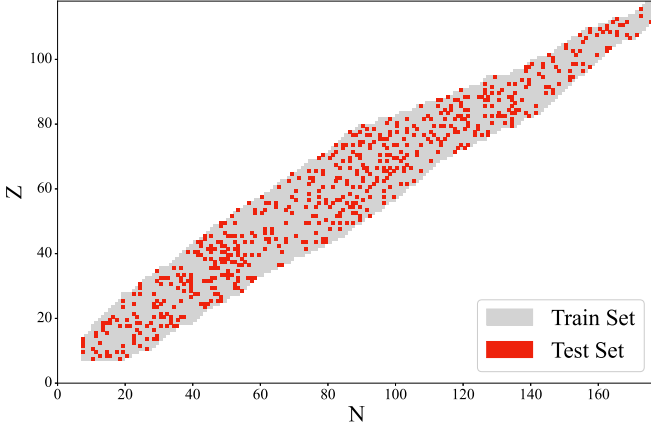
FIG. 2. The training set (gray circles), test set (red circles) used in the KANs. Both the training and test sets include nuclei from AME2020.

In each iteration, $k - 1$ parts are used for training, and the remaining part is used for validation. This process repeats k times, with each part serving as the validation set once. Averaging the performance across all iterations gives a reliable measure of the model's generalizability. In our calculations, k is set to 3, thereby allocating $1/3$ of the data for testing and the remaining $2/3$ for training in each cycle, with the initial randomization facilitated by the Fisher-Yates shuffle algorithm to ensure fairness and unpredictability in the distribution of data points across folds.

We carefully selected data attributes to ensure the framework's effectiveness and accuracy. These include the atomic number (Z), neutron number (N), and mass number (A). We also included features from previous research [13,20,31], such as pairing effects, shell closures, and asymmetry, which are crucial for modeling nuclear binding energies.

Pairing effects are captured using Z_{EO} and N_{EO}, which can only be either 1 or 0. Here, 1 indicates odd Z/N ratios, and 0 indicates even ratios. This distinction helps the model differentiate between various types of nuclei, each possessing unique binding energy characteristics. Additionally, the term $A^{2/3}$ represents the surface effect, reflecting that a nucleus's surface area scales with the two-thirds power of its mass number (A). This scaling is crucial for modeling the surface contributions to the nucleus's overall binding energy.

The isospin asymmetry, $(N - Z)$, measures the difference between neutrons and protons. We also consider nuclear magic numbers, with μ_Z and μ_N representing the excess numbers of protons and neutrons relative to the nearest closed shell. Key magic numbers include 8, 20, 28, 50, 82, 126, 184.

We also explore nuclear shells with Z_{shell} and N_{shell}, reflecting the orbital configurations of the last proton and neutron. These values range from 0 to 5, depending on proton or neutron counts within specific intervals. This helps the model capture the shell structure's influence on binding energy.

We evaluated various combinations of features to identify the most informative attributes for predicting nuclear binding energies. This approach helped us determine the optimal feature set for our KAN framework. Our network structure was shallow, consisting of a single hidden layer with a uniform

TABLE I. Feature space and KANs' structure.

Model	Feature	Structure
KAN-2	N, Z	[2,12,1]
KAN-4	N, Z, A, N-Z	[4,12,1]
KAN-9	N, Z, A, N-Z, $A^{2/3}$, Z_{EO}, N_{EO}, μ_Z, μ_N	[9,12,1]
KAN-11	N, Z, A, N-Z, $A^{2/3}$, Z_{EO}, N_{EO}, μ_Z, μ_N, Z_{shell}, N_{shell}	[11,12,1]

width of 12 neurons. The detailed features and network structure are presented in Table I. According to the requirements of Eq. (1), the number of nodes in the second layer should be at least $2n + 1$. Although this condition can be relaxed based on the work in Ref. [22], we have chosen to use the maximum value of $n + 1$ as the number of neurons in the hidden layer, which in our case is 12.

To improve model performance and interpretability while reducing overfitting, we used L1 regularization and cross-entropy loss as described in Ref. [22]. The loss function is defined as

$$\text{Loss} = \frac{(\text{BE}_{\text{pred}} - \text{BE}_{\text{exp}})^2}{N_p}$$
$$+ \lambda \left(\mu_1 \sum_{l=0}^{L-1} |\mathbf{\Phi}_l|_1 + \mu_2 \sum_{l=0}^{L-1} S(\mathbf{\Phi}_l) \right), \quad (2)$$

where $|\mathbf{\Phi}_l|_1$ is the L1 norm of the KAN layer, defined as the sum of the L1 norms of all activation functions $\phi_{i,j}$, including n_{out} outputs and n_{in} inputs. The L1 norm of an activation function, averaged over N_p inputs, is given by

$$|\phi|_1 \equiv \frac{1}{N_p} \sum_{s=1}^{N_p} |\phi(x^{(s)})|. \quad (3)$$

Thus, the L1 norm of a KAN layer is

$$|\mathbf{\Phi}|_1 \equiv \sum_{i=1}^{n_{\text{in}}} \sum_{j=1}^{n_{\text{out}}} |\phi_{i,j}|_1. \quad (4)$$

The entropy of a KAN layer is defined as

$$S(\mathbf{\Phi}) \equiv - \sum_{i=1}^{n_{\text{in}}} \sum_{j=1}^{n_{\text{out}}} \frac{|\phi_{i,j}|_1}{|\mathbf{\Phi}|_1} \log \left(\frac{|\phi_{i,j}|_1}{|\mathbf{\Phi}|_1} \right). \quad (5)$$

L1 regularization helps make the model simpler by adding a penalty based on the absolute values of the model's parameters. This leads to models with fewer nonzero parameters, making them easier to interpret and less likely to overfit. Cross-entropy loss ensures the model's predictions are well calibrated by measuring the difference between predicted and actual probability distributions.

Combining these techniques, we aim to create a robust and understandable model for predicting nuclear binding energies. We used k-fold cross-validation [27–29] to test different parts of the data, selected features carefully, and applied advanced regularization techniques to ensure our KAN framework is precise and generalizable. By exploring different features and
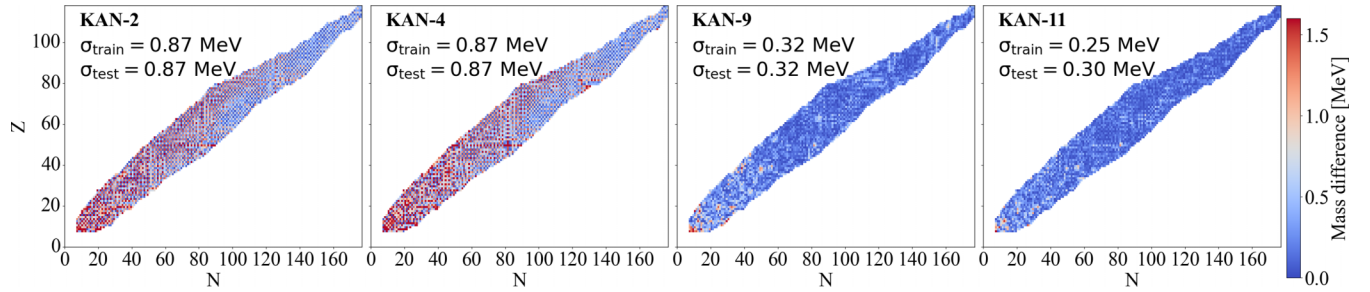
FIG. 3. The absolute value of binding energy differences between KAN predictions using different features (see Table I) and AME2020 [25]. The RMSEs for all the data are also provided.

III. RESULTS

Figure 3 shows the absolute differences in binding energy between predictions from various KAN models and the experimental data from AME2020 [25]. The feature sets used in these models are listed in Table I. Notably, the KAN-2 model, which uses only basic properties such as neutron number N and proton number Z, achieves a root mean square error (RMSE) of just 0.87 MeV. This performance surpasses that of many microscopic models [20], highlighting the effectiveness of the KAN approach even with minimal input features. When additional features are incorporated, the KAN-11 model reduces the RMSE even further to an impressive 0.26 MeV for the entire data set. Furthermore, the nearly identical RMSEs for both the training and test sets indicate the robustness achieved through the k-folding method and KANs.

As more features, such as isospin asymmetry and mass number, are included, the RMSE decreases, as illustrated in Fig. 3, progressing from KAN-2 to KAN-4. However, both models (KAN-2 and KAN-4) show a significant increase in deviation error as nuclei approach closed shells. This error distribution underscores the influence of shell effects on nuclear binding energy. By incorporating features related to shell structure, models KAN-9 and KAN-11 achieve lower RMSEs. Figure 3 demonstrates that these models perform better for medium and heavy nuclei compared to light nuclei. Despite efforts to enhance predictions by expanding the feature space, the challenge of modeling light nuclei persists.

While we acknowledge that increasing the total number of parameters can introduce complexity and potential issues, it's important to analyze this impact carefully. In our models, we selected 30 basis functions to represent each activation function. Although using more basis functions can better approximate the "true activation function," it also increases the model's parameter count and computational demands. The required number of basis functions largely depends on the properties of the selected basis functions and the behavior of the activation functions they aim to approximate. Therefore, selecting appropriate basis functions tailored to the specific problem and dataset can significantly reduce the number of parameters without compromising performance.

Previous research has explored various alternative basis functions beyond traditional B-spline curves. Researchers have experimented with sine functions, trainable adaptive fractional-orthogonal Jacobi functions, wavelets, polynomials, and rational functions based on Padé approximations and rational Jacobi functions [32–35]. For instance, the development of rational Kolmogorov-Arnold networks (rKANs) utilizes rational functions as basis functions, offering improved approximation capabilities while more effectively managing the parameter scale [33]. These alternative basis functions enhance the network's ability to model complex functions while optimizing computational efficiency. Therefore, in KANs, the specific choice of basis functions is not as critical; rather, the number of activation functions plays a more significant role. Therefore we set the same width for the hidden layer.

For all that, we can still compare the parameter scales of different KANs. To assess the parameter scale, we note that KAN-2 roughly contains 1080 free parameters; KAN-4 has about 1800; KAN-9 has 3000; and KAN-11 has 4320 free parameters. As mentioned earlier, for KAN-2 and KAN-4, we deliberately increased the width of their hidden layers to match the network scale of the more complex models. In practical applications, similar RMSE can be achieved with a narrower width. For example, for KAN-2, using architectures like [2,5,1] or [2,12,1], we achieved comparable RMSE on the dataset, and the parameter count of [2,5,1] is only 450— less than half of the original. Additionally, by applying L1 regularization, we can prune unimportant branches to obtain a smaller model, as demonstrated in Sec. III C.

We also compare our results with previous work. A recent application of mixture density networks (MDNs), a type of machine learning algorithm, resulted in errors of about 0.5–0.6 MeV compared to AME2016 when combined with physics-based features [15]. Gaussian process regression (GPR) achieved errors ranging from 0.14 to 0.96 MeV for the training set and 0.26 to 1.08 MeV for the test set compared to AME2020, using different physical features as reported in Ref. [20]. Support vector regression (SVR) reached errors from 0.39 to 2.55 MeV for the test set and from 0.23 to 2.40 MeV for the training set compared to AME2020, as shown in Ref. [20].

Meanwhile, Ref. [18] shows that Bayesian neural networks (BNNs) can achieve a RMSE of 84 keV. They used even-even nuclei as the foundation of their dataset and determined the
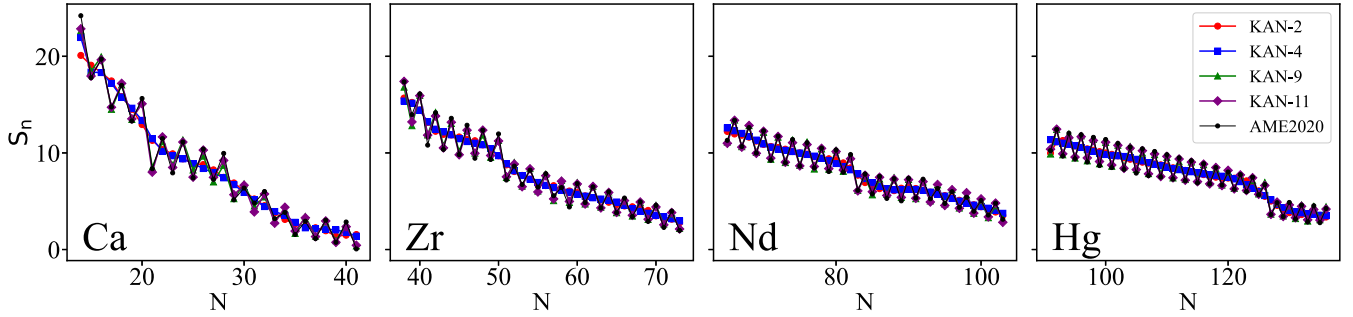
FIG. 4. Experimental single neutron separation energies for Ca, Zr, Nd, and Hg isotopic chains in comparison with the KAN predictions.

binding energies of surrounding nuclei by simultaneously predicting and combining the separation energies of neighboring nuclei. Compared to these models, KAN achieves similar RMSE, except for the BNN, which exhibits superior performance. This result has already surpassed the performance of most model-driven methods. For example, the corresponding RMSEs are 0.285, 0.559, and 0.576 MeV for the WS4 [31], FRDM2012 [36], and HFB-31 [37] models, respectively. Given that KAN is a novel technology, there is reason to believe that ongoing developments will further enhance its capabilities. As the technology matures, we can expect predictions to become even more accurate and precise, leading to broader applications in nuclear physics and machine learning.

Fairly comparing training times among different models is challenging due to variations in implementation, computational resources, and optimization strategies. However, as noted in Ref. [22], while KANs offer the advantage of reducing network size compared to traditional MLPs, MLPs tend to be more efficient in training networks of the same scale. However, KANs possess distinct properties, such as the ability to initiate training with a fewer number of basis functions for approximating activation functions, gradually refining them as needed. This capability allows us to start with a simpler model to capture the fundamental shape of activation functions and incrementally increase the basis functions to enhance detail. This approach simplifies the training process and may provide benefits in terms of interpretability and computational efficiency. Furthermore, recent studies [32–35] have demonstrated that various KAN variants are becoming increasingly efficient, potentially paving the way for more effective implementations in diverse applications.

Figure 4 compares the separation energies for isotopic chains of four different elements against the values in AME2020. The absence of pairing terms in KAN-2 and KAN-4 leads to much larger prediction RMSEs compared to KAN-9 and KAN-11, which include these terms. This difference is especially noticeable in the color variations representing the absolute differences in binding energy between the predictions from various KAN models and the experimental data from AME2020, as shown in Fig. 3. These variations are present in KAN-2 and KAN-4 but not in KAN-9 and KAN-11, highlighting the crucial role of pairing effects in detailing nuclear structure. The single-neutron separation energy curves in Fig. 4 further demonstrate the impact of

pairing. The difficulties faced by KAN-2 and KAN-4 models with abrupt changes suggest that adding features like N_{EO} and Z_{EO} could better model the pairing behavior in nuclear binding energies, leading to improved performance in KAN-9 and KAN-11.

However, small differences remain in the calcium (Ca) isotopic chain, as shown in Fig. 4. The model shows larger RMSEs for light nuclei compared to medium and heavy nuclei. Light nuclei are more affected by interactions among a few particles, while heavier nuclei are influenced by interactions among many particles. This difference makes it challenging to create a model that works well for all types of nuclei. It suggests that we need a more advanced KAN model to better understand the complex behavior of nuclear systems across different mass regions.

A. Extrapolation

Extrapolating beyond the training dataset presents a significant challenge for machine learning and deep learning models. To evaluate the extrapolation capability of our KANs, we initially focused on nuclei at the edges of the AME2020 dataset included in the test set—specifically, boundary isotopes and isotones such as $^{92}_{59}$As, $^{91}_{58}$As, $^{90}_{57}$As, and $^{215}_{135}$Hg, $^{216}_{135}$Tl, $^{217}_{135}$Pb. We observed that KANs could predict the binding energies of these nuclei with reasonable RMSEs. However, since these nuclei are very close to the data in the training set, a more thorough assessment of the KANs' ability to extrapolate was deemed necessary.

To accomplish this, we compared the predictions of our KANs with those from the finite-range droplet model (FRDM2012) [36]. Figure 5 showcases the mass excess predictions made by KANs with various feature sets alongside the results from FRDM2012. This dataset encompasses 9138 nuclei, ranging from ^{16}O up to a mass number $A = 339$, significantly extending beyond the scope of AME2020.

Our observations revealed that the KAN-4 and KAN-11 models demonstrated good agreement with FRDM2012 across most regions, with many differences between these models and FRDM2012 being within 50 MeV. Considering we are dealing with nuclei far from the known mass boundaries, such discrepancies between different models are deemed acceptable. The KAN-2 model, however, exhibited limited extrapolation capability, which is understandable
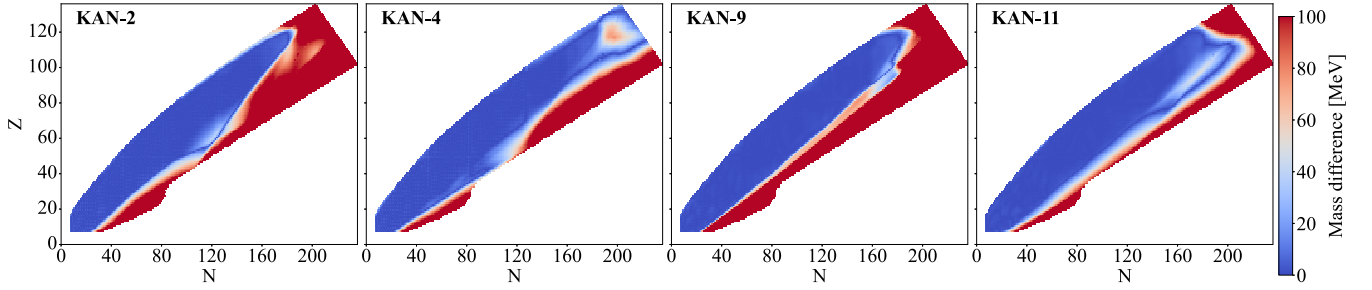
FIG. 5. The absolute value of mass excess differences between KAN predictions using different features (see Table I) and FRDM12's prediction [36].

given its minimal feature set. Interestingly, the KAN-9 model showed signs of overfitting to a certain extent.

Furthermore, in regions significantly distant from the training data, our KAN models exhibit larger deviations from the results predicted by the physically motivated FRDM2012 model. This discrepancy can be attributed to two main factors. First, the divergence among different models tends to increase when extrapolating far beyond the known data regions. This phenomenon is also observed in Refs. [17,20], where differences among various density functional models and machine learning models are noted. Second, the use of B-splines as basis functions for expanding the activation function means the model's behavior beyond the training set heavily relies on the edge behavior of the B-spline curves. Essentially, the extrapolation is influenced more by the polynomial behavior at the boundaries rather than by the intended activation functions.

A potential solution to improve extrapolation capabilities could involve employing symbolic regression to replace the polynomial expansion form. Functions derived from symbolic regression may inherently possess better extrapolation capabilities due to their flexibility and the ability to capture underlying patterns more effectively.

On the other hand, as depicted in Fig. 6, when predicting isotopic chains extending ten neutrons beyond the existing dataset, the KANs' predictions closely align with those from FRDM2012. This suggests that the KAN networks have reliable extrapolation abilities in regions not too distant from the training data, demonstrating stable and trustworthy performance. It is also evident that KAN performs better on

heavy nuclei than on light nuclei, consistent with the observations made in Figs. 3 and 4.

B. Feature analysis

In Fig. 7 we show the structure of the trained KAN. To clarify node connections, we adjust the transparency of lines based on the mean value of the activation function $\phi_{l,i,j}$ for each connection, represented by $\tanh(\beta A_{l,i,j})$, where $A_{l,i,j}$ is the mean activation. This technique highlights each variable's influence on the final output, aiding in understanding the network's behavior. Additionally, Fig. 7(b) illustrates the $L1$ norm for each connection, providing insights into the importance of each link and identifying significant contributors to the model's performance. This dual visualization of mean activation and $L1$ norm offers a comprehensive view of the network's workings and the relative importance of different features and connections.

Our analysis indicates that the neutron number N exerts the largest influence on the model's predictions, as evidenced by its highest $L1$ norm among the variables. The nucleon number A and proton number Z also significantly impact the model, with their $L1$ norms being large and comparable to that of N. This suggests that N, A, and Z are the primary factors influencing nuclear properties in our model.

The asymmetry term $N - Z$ and the surface term $A^{2/3}$ exhibit moderate effects, highlighting their relevance but lesser dominance compared to A, N, and Z. These terms are crucial for capturing the nuances of nuclear structure, though their impact is not as pronounced as the primary variables.
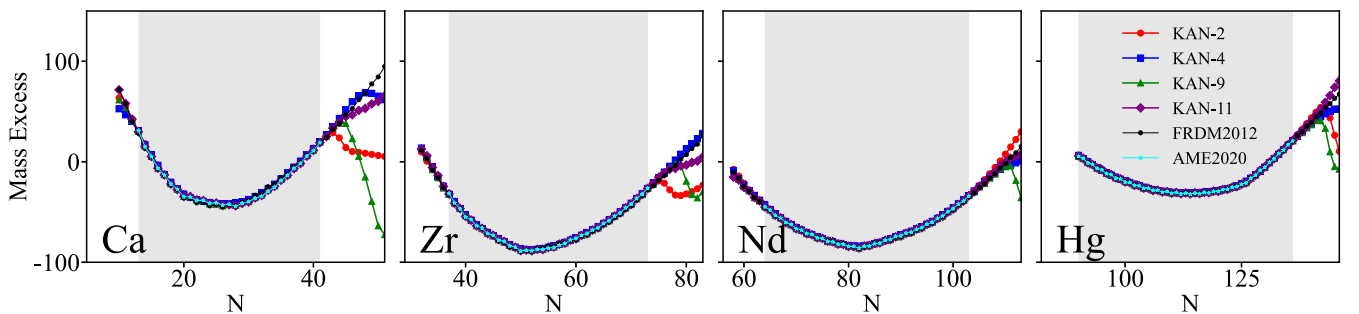

FIG. 6. Mass excess for Ca, Zr, Nd, and Hg isotopic chains compared with KANs predictions, FRDM12 (black points) [36], and experimental data from AME2020 (cyan stars) [25].
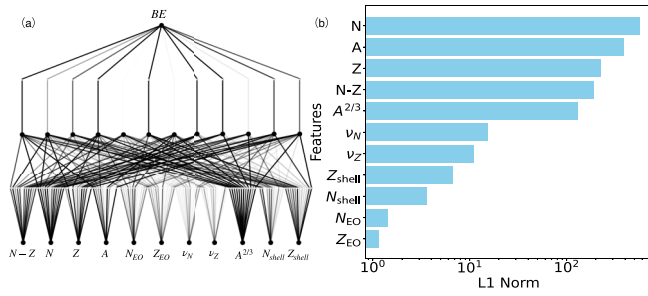
FIG. 7. Visualization of the trained network structure in the panel (a). The transparency of the lines connecting nodes is proportional to $\tanh(\beta A_{l,i,j})$, where $A_{l,i,j}$ is the mean activation, enhancing the visibility of connections. In panel (b), the $L1$ norm for each connection is shown on a logarithmic scale, highlighting the significance of each link.
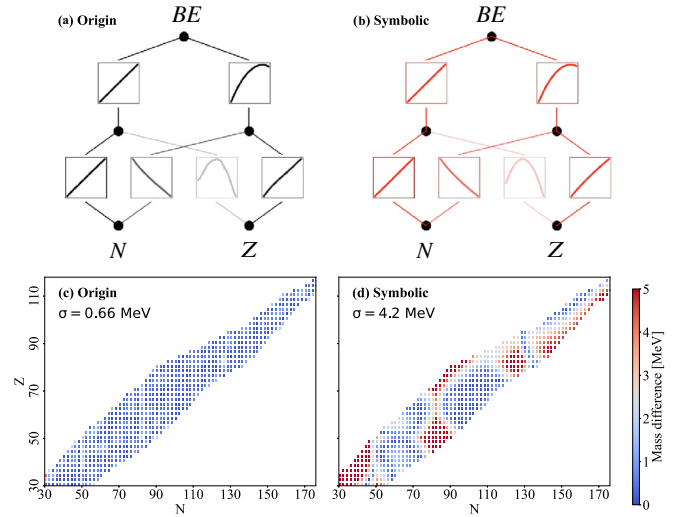

FIG. 8. Panel (a) shows the original trained network structure, while panel (b) shows the symbolic network structure. Panels (c) and (d) illustrate the absolute value of binding energy differences between predictions and data from AME2020 [25].

Variables associated with nuclear shell structure and the pairing term demonstrate minimal direct impact on the overall results. Despite their subtle influence, these factors are essential for accurate predictions of nuclear binding energies, especially in astrophysics and nucleosynthesis, where high accuracy and precision are paramount. Figures 3 and 4 show that incorporating these terms, although they contribute less directly, is vital for modeling.

These findings align with our calculations and are supported by results from other studies [13,15,20,31,38], reinforcing the validity of our approach and the robustness of our model. This analysis underlines the primary drivers of nuclear binding energy and emphasizes the importance of considering all relevant factors, even those with less apparent impact, to accurately capture the complexity of nuclear interactions.

It is important to note, however, that this analysis can only provide approximate relationships between the variables. Due to the inherent randomness in the training process—such as the random initialization of model weights and the stochastic nature of optimization algorithms—the exact values of the $L1$ norms may vary across different training sessions. This means that while we can identify general patterns and trends, establishing absolute relationships between the variables based on a single instance of training is not possible.

C. Symbolic regression

One of the most intriguing aspects of Kolmogorov-Arnold networks (KANs) is their ability to directly derive symbolic expressions from data, which enhances both the interpretability and transparency of the predictive model. Unlike traditional neural networks that employ fixed activation functions—potentially obscuring underlying patterns—KANs learn activation functions tailored to the data. These can initially be represented using B-splines, which are smooth, piecewise-defined functions.

To extract analytical formulas for nuclear binding energy from these learned activation functions, we establish a function library composed of elementary functions and basic functional forms, such as polynomials, exponentials, and logarithms. By utilizing the least-squares method, we adjust the

coefficients of these functions to best match the learned activation functions, effectively performing symbolic regression. Afterward, by combining each activation function through Eq. (1), we obtain the final expression, with specific details available in Ref. [22].

To simplify the model and reduce complexity introduced by periodic functions (like sine functions that may be influenced by pairing terms), we focus on even-even nuclei. This selection minimizes the influence of pairing effects on the activation functions. Additionally, we limit our study to medium to heavy nuclei, with proton and neutron numbers $Z \geqslant 30$ and $N \geqslant 30$, as these nuclei exhibit more uniform behavior conducive to symbolic modeling.

We initially trained a KAN with a straightforward architecture: two input neurons, one hidden layer with five neurons, and one output neuron. To enhance network sparsity and maintain performance while increasing the regularization coefficient, we pruned connections with minimal impact on the output. This pruning simplified the network to one with two input neurons, one hidden layer containing two neurons, and one output neuron. Remarkably, the final model achieved a RMSE of 0.66 MeV. With this streamlined network, we proceeded to perform symbolic regression on the learned activation functions.

It's important to acknowledge that approximating each activation function with elementary mathematical functions inevitably introduces some error. These approximation errors can accumulate and propagate through the network, potentially leading to larger discrepancies in the final predictions compared to the original KAN model. Consequently, the symbolic regression results may exhibit greater errors relative to the initial model. Nevertheless, the derived analytical expressions provide valuable insights into the underlying physical laws governing nuclear binding energies.

As illustrated in Fig. 8, transitioning from the original network's activation functions [the black line in Fig. 8(a)] to

TABLE II. Parameter and its value for Eq. (6).

Parameter	Value	Parameter	Value
A_{vol} (MeV)	9.87917	A_N (MeV$^{1/2}$)	0.53670
A_{vol} (MeV)	−1.56157	A_Z (MeV$^{1/2}$)	2.93091
A_{coul} (MeV)	0.060169	n	−1.06419
C_1 (MeV)	13.08641	z	6.26334
C_2 (MeV$^{1/2}$)	11.76810		

the symbolic network's activation functions [the red line in Fig. 8(b)] results in some loss of detail. These lost details propagate through the network, leading to varying degrees of error in the predictions. A comparison of the differences between the original and symbolic models against experimental data [Figs. 8(c) and 8(d)] reveals underlying patterns: near magic numbers 50, 82, and 126, the errors in the symbolic results are notably larger than those in the original results. This suggests that the nuances not fully captured by the symbolic approximations in the activation functions reflect structural information about the nuclei—information that is crucial near these magic numbers due to shell closures and associated nuclear structure effects.

Despite not including variables explicitly related to shell structure in our input features, the KAN demonstrated a robust capability to identify and learn these features directly from the data. This ability underscores the strength of KANs in capturing complex physical phenomena and validates the effectiveness of our symbolic regression approach in revealing and confirming underlying physical laws, even when working with simplified models and limited input variables.

Below is the expression obtained through symbolic regression. After simplification, it can be presented in the following form, and the parameters are listed in Table II:

$$BE(N, Z) = A_{vol}A - A_{sym}(N - Z) - A_{coul}Z^2$$
$$- [A_N(N + n)^{2/3} - A_Z(Z + z)^{2/3} + C_2]^2 - C_1.$$
(6)

This expression for the binding energy achieves an RMSE of 4.2 MeV for these even-even nuclei. Given that the binding energies for nuclei in this range lie between 500 and 2100 MeV, this suggests that the formula retains less than 1% error.

When compared to the classical liquid drop model and the Bethe-Weizsäcker formula, the initial three terms in the equation align with the volume, symmetry energy, and Coulomb terms, respectively. The fourth term involves the square of the difference between the cube roots of the neutron and proton numbers, reflecting the surface energy effects of the nucleus. Here, A_N and A_Z are the surface energy coefficients for neutrons and protons, respectively, and n and z are adjustments to the neutron and proton numbers to capture more subtle nuclear structural effects. C_2, as a translation constant, adjusts the overall level of the surface energy term. The final constant, C_1, shifts the entire energy expression to optimize the model fit.

It is also worth noting that an underlying relationship exists between N and Z due to the stability and other properties of nuclei. Classical distributions of nuclei often adhere to certain N/Z ratios, as suggested by the Bethe-Weizsäcker formula. This implies that N and Z are not entirely independent variables and can, to some extent, be interchanged. Therefore, the derived analytical expression may not be unique.

As observed in Fig. 9 and consistent with our previous analysis, the application of symbolic regression enhances the stability of extrapolation. For the isotopic chains of Zr, Nd, Hg, and Fm, the predictions obtained through symbolic regression (depicted as blue squares) exhibit greater stability compared to those from the original KAN approach (represented by red dots). This indicates that symbolic regression provides corrective effects of varying magnitudes at the boundaries of the data, thereby improving the model's predictive performance in these regions.

However, it is important to acknowledge that symbolic regression cannot entirely eliminate the issues associated with extrapolating beyond the range of the available data. Such extrapolation can lead to excessively weak constraints on the predicted results, resulting in potential model failures. Therefore, while symbolic regression contributes to improved extrapolation stability, it does not fully overcome the inherent limitations of extrapolation in machine learning models. Care must still be taken when making predictions outside the scope of the training data to ensure the reliability of the results.

IV. SUMMARY AND CONCLUSIONS

In this study, we have demonstrated the effectiveness of KANs in predicting nuclear binding energies. By leveraging
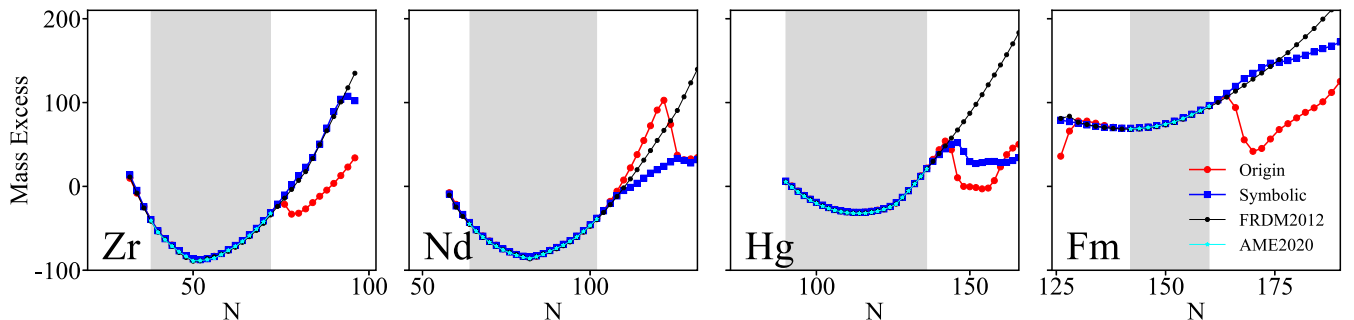

FIG. 9. Mass excess for Zr, Nd, Hg, and Fm isotopic chains compared with KANs predictions, the results from symbolic regression, FRDM12 (black points) [36], and experimental data from AME2020 (cyan stars) [25].

the capabilities of KANs to decompose complex multivariate functions into univariate components, we have shown that even with minimal input features, KANs can achieve significantly lower RMSE, surpassing many traditional physics-based models. The incorporation of additional features related to isospin asymmetry, shell structure, and pairing effects further reduces the RMSE of the predictions, emphasizing the importance of these factors in modeling nuclear binding energies.

Our results show that KANs can effectively capture the underlying physical relationships in nuclear data, achieving a RMSE as low as 0.26 MeV for the entire dataset when using an expanded set of features. The KAN models exhibit robust performance across different regions of the nuclear chart, particularly for medium and heavy nuclei. However, challenges remain in modeling light nuclei, suggesting the need for more advanced models or additional features to capture the complexities associated with few-body dynamics.

The symbolic regression analysis conducted in this work yielded an analytical expression for nuclear binding energy that aligns well with established models like the Bethe-Weizsäcker formula. This expression incorporates key terms such as volume energy, symmetry energy, Coulomb energy, and surface energy, providing direct insight into the functional dependencies within the nuclear data. While the symbolic model exhibited marginally higher RMSE compared to the original KAN model, it offers the advantage of improved interpretability and potential for deeper theoretical understanding.

We also explored the extrapolation capabilities of KANs by comparing their predictions with those from the finite-range droplet model (FRDM2012) [36] for nuclei beyond the known mass boundaries. The KAN models demonstrated reasonable agreement with FRDM2012 for nuclei not too distant from the training data, indicating stable and trustworthy performance in extrapolation within similar regions. However, larger deviations were observed when predicting nuclei far beyond the training set, highlighting inherent limitations in extrapo-lation and the need for caution when predicting properties of nuclei in uncharted regions.

Overall, our study highlights the potential of KANs in nuclear physics applications, combining data-driven approaches with theoretical insights to enhance predictive capabilities. The ability of KANs to balance the error with model simplicity and interpretability makes them a promising tool for exploring complex systems where traditional methods may face challenges.

Future work will focus on extending the application of KANs to better model light nuclei and refine the network architectures for improved performance. Additionally, further exploration of symbolic regression techniques within the KAN framework could lead to new theoretical developments and more accurate models in nuclear physics and other fields involving complex multivariate relationships.

An important direction for future research is the incorporation of uncertainty quantification, as demonstrated in recent advancements with KANs [39]. By employing a Bayesian approach and enhancing computational efficiency with methods like those proposed for ReLU-KANs, both epistemic and aleatoric uncertainties can be effectively quantified [39]. This not only broadens the applicability of KANs to nuclear physics by ensuring accurate identification of functional dependencies in stochastic environments, but also enhances the reliability and robustness of predictions. Implementing these methods could pave the way for innovative applications in nuclear physics, providing insights into the model's uncertainty characteristics and revealing areas for improvement.

[1] H.-W. Hammer, S. König, and U. van Kolck, Rev. Mod. Phys. **92**, 025004 (2020).

[2] C. v. Weizsäcker, Z. Angew. Phys. **96**, 431 (1935).

[3] M. Rosenbusch, P. Ascher, D. Atanasov, C. Barbieri, D. Beck, K. Blaum, C. Borgmann, M. Breitenfeldt, R. B. Cakirli, A. Cipollone, S. George, F. Herfurth, M. Kowalska, S. Kreim, D. Lunney, V. Manea, P. Navrátil, D. Neidherr, L. Schweikhard, V. Somà *et al.*, Phys. Rev. Lett. **114**, 202501 (2015).

[4] Q. Mo, M. Liu, and N. Wang, Phys. Rev. C **90**, 024320 (2014).

[5] D. Lunney, J. M. Pearson, and C. Thibault, Rev. Mod. Phys. **75**, 1021 (2003).

[6] M. Arnould, S. Goriely, and K. Takahashi, Phys. Rep. **450**, 97 (2007).

[7] H. Schatz and W.-J. Ong, Astrophys. J. **844**, 139 (2017).

[8] H. A. Bethe and R. F. Bacher, Rev. Mod. Phys. **8**, 82 (1936).

[9] W. Myers and W. Swiatecki, Nucl. Phys. A **601**, 141 (1996).

[10] S. Goriely, S. Hilaire, M. Girod, and S. Péru, Phys. Rev. Lett. **102**, 242501 (2009).

[11] M. Hadizadeh, M. Radin, and K. Mohseni, Sci. Rep. **10**, 1949 (2020).

[12] E. Epelbaum, H.-W. Hammer, and U.-G. Meißner, Rev. Mod. Phys. **81**, 1773 (2009).

[13] L.-X. Zeng, Y.-Y. Yin, X.-X. Dong, and L.-S. Geng, Phys. Rev. C **109**, 034318 (2024).

[14] C.-Q. Li, C.-N. Tong, H.-J. Du, and L.-G. Pang, Phys. Rev. C **105**, 064306 (2022).

[15] A. E. Lovell, A. T. Mohan, T. M. Sprouse, and M. R. Mumpower, Phys. Rev. C **106**, 014305 (2022).

[16] Y. Liu, C. Su, J. Liu, P. Danielewicz, C. Xu, and Z. Ren, Phys. Rev. C **104**, 014315 (2021).

[17] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, Phys. Rev. C **98**, 034318 (2018).

[18] Z. M. Niu and H. Z. Liang, Phys. Rev. C **106**, L021303 (2022).

[19] Z.-Y. Yuan, D. Bai, Z. Wang, and Z.-Z. Ren, Nucl. Sci. Tech. **35**, 105 (2024).

[20] E. Yüksel, D. Soydaner, and H. Bahtiyar, Phys. Rev. C **109**, 064322 (2024).

[21] J. M. Munoz, S. M. Udrescu, and R. F. G. Ruiz, arXiv:2404.11477.

[22] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, *The Thirteenth International Conference on Learning Representations* (2025).

[23] A. N. Kolmogorov, Dokl. Akad. Nauk. USSR **114**, 953 (1957).

[24] V. I. Arnol'd, On the representation of functions of several variables as a superposition of functions of a smaller number of variables, in *Collected Works Representations of Functions, Celestial Mechanics and KAM Theory, 1957–1965*, edited by A. B. Givental, B. A. Khesin, J. E. Marsden, A. N. Varchenko, V. A. Vassiliev, O. Y. Viro, and V. M. Zakalyukin (Springer, Berlin, Heidelberg, 2009), pp. 25–46.

[25] M. Wang, W. Huang, F. Kondev, G. Audi, and S. Naimi, Chin. Phys. C **45**, 030003 (2021).

[26] A. Menditto, M. Patriarca, and B. Magnusson, Accredit. Qual. Assur. **12**, 45 (2007).

[27] D. M. Allen, Technometrics **16**, 125 (1974).

[28] M. Stone, J. R. Stat. Soc. B **36**, 111 (2018).

[29] M. Stone, J. R. Stat. Soc. B **39**, 44 (2018).

[30] R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver & Boyd, Edinburgh, 1953), pp. 26–27, 4th ed.

[31] N. Wang, M. Liu, X. Wu, and J. Meng, Phys. Lett. B **734**, 215 (2014).

[32] A. A. Aghaei, arXiv:2406.07456.

[33] A. A. Aghaei, arXiv:2406.14495.

[34] X. Warin, arXiv:2410.03801.

[35] E. A. F. Reinhardt, P. R. Dinesh, and S. Gleyzer, Front. Artif. Intell. **7**, 1462952 (2025).

[36] P. Möller, A. Sierk, T. Ichikawa, and H. Sagawa, At. Data Nucl. Data Tables **109-110**, 1 (2016).

[37] S. Goriely, N. Chamel, and J. M. Pearson, Phys. Rev. C **93**, 034337 (2016).

[38] M. R. Mumpower, T. M. Sprouse, A. E. Lovell, and A. T. Mohan, Phys. Rev. C **106**, L021301 (2022).

[39] J. Giroux and C. Fanelli, arXiv:2410.01687.