Physically interpretable machine learning for nuclear masses

M. R. Mumpower ●,* T. M. Sprouse ●, and A. E. Lovell ●
Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

A. T. Mohan ●
Computer, Computational and Statistical Sciences Division,
Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

We present an approach to modeling the ground-state mass of atomic nuclei based directly on a probabilistic neural network constrained by relevant physics. Our physically interpretable machine learning (PIML) approach incorporates knowledge of physics by using a physically motivated feature space in addition to a soft physics constraint that is implemented as a penalty to the loss function. We train our PIML model on a random set of approximately 20% of the atomic mass evaluation (AME) and predict the remaining 80%. The success of our methodology is exhibited by a $\sigma_{rms} \approx 186$ keV match to data for the training set and $\sigma_{rms} \approx 316$ keV for the entire AME with $Z \geqslant 20$. We show that our general methodology can be interpreted using feature importance.

Introduction. The minimal energy required to break up a nucleus into its constituent nucleons is one of the fundamental properties of an atomic nucleus. This quantity, which is equivalent to the mass, features prominently as an input for the theoretical prediction of a number of nuclear properties which are important for both scientific and technological applications [1–3]. This effect is perhaps most apparent in the important role that masses play in predicting the reaction and decay properties of atomic nuclei [4,5]. Masses also serve as critical inputs for the study of astrophysical phenomena from influencing the composition of neutron star crusts to impacting heavy element synthesis and its potential observable consequences [6–10].

The many-body Hamiltonian that describes atomic nuclei is exceedingly complex and remains unknown, therefore, the lowest-energy state cannot be calculated directly from first principles for heavy nuclei. This state of affairs has led to the development of many theoretical descriptions of atomic masses, including semiclassical approaches [11], microscopic approaches [12], and more recently, models enhanced by considering improvements to model discrepancies with data [13,14]. An inherent limitation in contemporary modeling is that the model itself remains fixed with optimization focused on parameters. This can be overcome with application of machine learning (ML) algorithms in which the model itself is optimized [15].

A wide range of approaches that utilize ML methods have been adopted by the nuclear physics community over the past several years [16]. Artificial neural networks have shown great promise as an extrapolation tool for *ab initio* nuclear theory of light nuclei [17,18]. Physics-informed ML using Bayesian techniques, such as maximum entropy have been used to inform the inverse of the Laplace transform required by electroweak response functions [19]. Finite-size effects in many-body physics have utilized neural networks to extrapolate the unitary gas to the thermodynamic limit at zero range [20]. Future applications, particularly, with artificial intelligence methods, such as natural language processing, will provide enormous assistance, notably with the extraction of key information and correction of compilation errors [21] in nuclear data evaluations [22].

Neural networks have also been used to study ground-state mass predictions. To date the most common application of ML methods to masses involve performing corrections to existing theoretical approaches [13,23–26]. These methods focus on training with residuals rather than directly with data. This results in a correction function which may be used to improve the match to existing data and may also be used for extrapolation [27]. Direct predictions of masses with a neutral networks are also beginning to be explored in the literature [28]. This includes a recent method that seeks to predict masses with a double network architecture [29]; the first network predicts the mass, and the second network performs error correction. Another use of ML mass predictions in the literature involves calculating relevant Q values for nuclear reactions [30]. The continual accumulation of accurate nuclear data will further empower these methods [31].

In this Letter we present an approach to modeling masses directly from a ML model constrained by physics. Our "physically interpretable machine learning" or (PIML) approach

---

*mumpower@lanl.gov; https://www.matthewmumpower.com.

builds physically meaningful feature spaces and applies soft constraints to ensure relevant physics is being obeyed. In contrast to all existing work, we show for the first time that an ML-based model utilizing a soft constraint is capable of learning masses without reference to any underlying theoretical model. Aside from raw nuclear data, we only posit relevant physical input features and the existence of major shell closures. Our methodology can be generalized to any problem in which physical constraints may need to be applied to a machine learned model. In a drastic improvement to our previous work [32], we train a probabilistic network on a fraction of available data and predict the vast majority of masses for thousands of nuclear species measured to date. We achieve outstanding model accuracy and retain predictive power when extrapolating. We show that our model can be interpreted using a standard measure for feature importance, and this interpretation fits within the context of the well-established picture of atomic nuclei.

Methods. We use a probabilistic ML technique, the mixture density network (MDN) [33]. Our probabilistic network is built on the PYTORCH [34] framework and can be run on either CPU or GPU architectures. This type of modeling has been shown to be successful in describing nuclear properties whereas providing well-quantified uncertainties even for complex quantities, such as fission distributions [35].

Lovell *et al.* [32] reported that a combination of macroscopic and microscopic features is suitable for describing masses across the chart of nuclides. Based on this previous analysis, we use eight features: the proton number (Z), the neutron number (N), the mass number (A), the odd-even nature of protons (Z_{eo}), the odd-even nature of neutrons (N_{eo}), the valence number of protons as measured from the nearest closed shell (V_p), the valence number of neutrons as measured from the nearest closed shell (V_n), and a measure of isospin asymmetry ($P_{asym} = \frac{N-Z}{A}$).

The last five features inform the model on quantum-mechanical effects. Pairing effects manifest from the inclusion of the Z_{eo} and N_{eo} terms which are binary, taking the value of 0 or 1. Valence terms characterize the counting of particles (or holes) between major closed neutron and proton shells. As the valence number increases up to the midshell, more complex excitations, including collective modes may appear [36]. The success of this picture can be related to nuclear promiscuity (a measure of the strength of proton-neutron interactions per valence nucleon) [37]. The final feature informs the model about the Pauli exclusion principle.

We take as our training set a random selection of the masses of 450 nuclei in the atomic mass evaluation 2016 (AME2016) [38] with $Z \geqslant 20$. The same set of 450 nuclei is fixed throughout training and does not change. The match to this data is computed with a logarithm loss function that we denote by $\mathcal{L}_1$ (see Ref. [32]).

In addition to the physics-based feature space, we seek to encode physical constraints into model training. For this Letter, we chose to employ one such possibility, the Garvey-Kelson (GK) relations [39]. This well-known series of formulas involves a judicious choice of mass differences of neighboring nuclei that minimizes the interactions between nucleons to first order, resulting in particular linear

combinations that strategically sum to zero. In this Letter we consider the following GK relations. If $N \geqslant Z$, we minimize the mass difference,

$$
\begin{aligned}
M(Z-2, N+2) - M(Z, N) + M(Z-1, N) \\
- M(Z-2, N+1) + M(Z, N+1) \\
- M(Z-1, N+2) \approx 0,
\end{aligned}
\tag{1}
$$

and for $N < Z$,

$$
\begin{aligned}
M(Z+2, N-2) - M(Z, N) + M(Z, N-1) \\
- M(Z+1, N-2) + M(Z+1, N) \\
- M(Z+2, N-1) \approx 0,
\end{aligned}
\tag{2}
$$

as in Ref. [40].

We implement these two equations as a soft constraint, a second loss function $\mathcal{L}_2$ in our training. This additional loss function is calculated *using only the mass predictions of the ML model*; no experimental data enter into the calculation of $\mathcal{L}_2$. The value of $\mathcal{L}_2$ is obtained by calculating the absolute value of the total sum of the GK relations for each nucleus in the set that defines the AME. This loss serves as a penalty for model solutions that do not obey well-established physical law. We, therefore, seek parameter spaces that minimize the value of $\mathcal{L}_2$ in the multiobjective optimization procedure that follows. Note that because our implementation is enacted as a soft constraint, training may ensue that temporarily increases $\mathcal{L}_2$ in pursuit of the global minimum. We revisit this important point shortly.

The model hyperparameters are as follows: The number of hidden layers is six, the number of hidden nodes is eight, the number of Gaussian ad mixtures is one (as mass is a scalar quantity), the weight of the physics constraint is $\lambda_{phys} = 1$, and we implement the Adam optimizer with learning rate 0.0002 [41]. To avoid overfitting, we implement regularization with a weight decay set to 0.01. These hyperparameters were determined from a select set of runs where the values were varied.

The weight of the physics constraint, $\lambda_{phys} = 1$ is especially noteworthy. We found that if the physics constraint was weighted too heavily (large values of λ_{phys}), training often failed as sharp cusps were encountered in the evolution of the total loss function which was prohibitive to optimization. As λ_{phys} tends to zero, the physics constraint becomes less influential on training and we return to the previous results of Ref. [32].

In training we seek to minimize the total logarithm loss that consists of a sum of the loss for the match to our training set as well as the physical constraint: $\mathcal{L}_{total} = \mathcal{L}_1 + \lambda_{phys}\mathcal{L}_2$. Each training epoch attempts to improve the total logarithm loss function with respect to previous solutions. We allow this process to continue for roughly 10^7 epochs.

The logarithm losses for a sample training run used in this Letter is shown in Fig. 1. The loss with respect to data $\mathcal{L}_1$ decreases monotonically by virtue of the minimization algorithm. The loss with respect to physics $\mathcal{L}_2$, however, exhibits a complex and highly nonlinear structure. This results in a total loss function that displays similarly complex behavior, e.g., briefly rising (ca. epoch no. 10^5), before establishing a global minimum for the entire run.
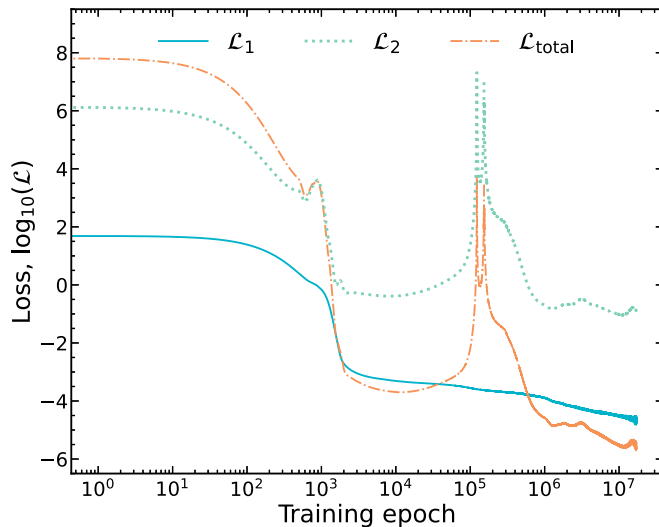
FIG. 1. The logarithm loss as a function of training epoch for the match to data $\mathcal{L}_1$ physics constraint $\mathcal{L}_2$, and the sum of the two $\mathcal{L}_{\text{total}} = \mathcal{L}_1 + \lambda_{\text{phys}}\mathcal{L}_2$. The learning rate is 0.0002.

Figure 1 demonstrates this very general but often overlooked attribute of optimization with multiple constraints [42,43]. If we consider epoch numbers $\approx 1-8 \times 10^5$, for example, we clearly see that whereas training based on fit to data *alone* is generally quite good (and improving), the same fit's ability to satisfy basic physical requirements is quite poor by comparison. Global minima across all epochs, e.g., in the case of a different training set, may not necessarily be the final epoch at which a stop condition is reached (although the two points do coincide by happenstance in Fig. 1). Consequently, it is important that past work that has applied ML in physics *without implementing or considering physically motivated constraints* should be approached with caution.

Results. In Fig. 2, we compare the PIML model predictions (blue) against our training data (AME2016 atomic masses) over a range of neodymium isotopes with existing measured data. The entirety of the AME dataset lies within the 3σ uncertainty interval with a majority (all but ≈ 3) data points lying within a 1σ or 2σ interval. The overall excellent agreement with AME data shown here is indicative of the results across the entire chart of nuclides. We report a root-mean-square error σ_{rms} for the PIML model to be $\sigma_{\text{rms}} \approx 186$ keV for the training set (20% coverage of AME2016 data) and $\sigma_{\text{rms}} \approx 316$ keV for the entirety of AME2016 with $Z \geqslant 20$. We retain our predictive capability ($\sigma_{\text{rms}} \approx 336$ keV) when comparing to the latest 2020 release of the evaluation [44]. These results are competitive with global mass models available today.

The green shaded band in Fig. 2 indicates model variations in training *without* inclusion of the GK relations. The upper and lower values of this band are calculated from the maximum and minimum of the mean values of the predictions of our previous models (M6, M8, M10, and M12) [32]. Whereas the green band captures the general trend in masses along this isotopic chain, training with the GK relations (this Letter, blue) enables a more refined prediction of features found in the mass surface. The difference shown for this isotopic chain is representative of other isotopic chains across the chart of
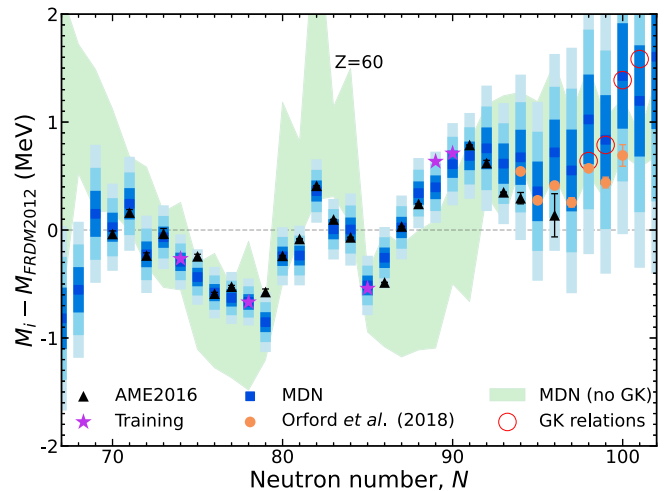

FIG. 2. Predictions of masses along the neodymium ($Z = 60$) isotopic chain using our PIML model. The PIML average and $1-3\sigma$ uncertainties are shown in progressively lighter shading along with AME2016 [38] and recent data from Ref. [45]. Nuclei used for training indicated by stars. Masses are plotted relative to FRDM2012. Also shown is a range of predictions without the GK relations (green band) and a recursive application of the GK relations (open red circles); see the text for details.

nuclides. This result underscores the importance of referencing physics constraints in training as it has a substantial impact on model predictability. Note that the green band designates a deviation in the maximum and minimum values alone; the variation in the associated individual uncertainties is not shown here.

A recursive application of the GK relations is shown by open red circles in Fig. 2. These relations perform exceedingly well close to known mass measurements where there is only a single unknown mass [46]. To extrapolate further, recursion is used [47]. However, the uncertainty quickly compounds upon repeated iteration limiting the predictive capability of this approach [48]. This is in contrast to the present Letter which seeks to fulfill the constraints of the above equations without repeated iteration.

Of primary interest to the nuclear physics community is not just the ability to model the properties of known nuclei (here, *masses*), but also the ability to apply these same models to predict the properties which currently cannot be produced or otherwise studied in laboratory settings. This has, to date, been a particularly difficult situation for nuclear theory, because whereas, e.g., microscopic, macroscopic-microscopic, or phenomenological approaches to modeling atomic nuclei may be applied to nuclei well outside the range of those used for parameter calibrating, it has proven challenging to develop a robust picture of the overall uncertainty in the extrapolated predictions. Recent efforts have included the application of Gaussian processes and Bayesian methods to better quantify the uncertainties of model parameters, particularly, those of microscopic nuclear models (see Refs. [14,49,50] and references therein).

Figure 2 gives some insight into how we may proceed with regard to understanding both the quality of PIML

extrapolations beyond available data as well as how these predictions can assist in our understanding of *total* (statistical, systematic, and/or model) uncertainties of nuclear mass predictions. In particular we note that the results of our calculations were derived from both training and testing against AME2016 data. In the time that has since passed, there have been a number of experiments which have provided data that extend beyond the limits of this dataset, and, in particular, the results of Ref. [45] extended the measurements of the neodymium isotopic chain shown in Fig. 2 towards a series of more neutron-rich isotopes up to atomic mass number $A = 160$.

On the question of extrapolation, we see that PIML predictions (which were constructed with no information concerning the data of Orford *et al.* [45]) would appear to generally follow the trends in nuclear mass suggested by the additional measurements, and indeed lying well within the 2σ-uncertainty intervals. This suggests that the PIML model (or other models founded on similar principles) may reasonably extrapolate towards the more exotic short-lived nuclei without any grotesque violation of basic nuclear physics principles, consistent with the design goals of our approach as laid out in the Methods section.

Furthermore, the PIML approach naturally provides thoroughly robust estimates with respect to the uncertainties pointwise for each individual prediction, i.e., each individual mass prediction is associated with its own uniquely inferred uncertainty. Indeed, the uncertainties shown in Fig. 2 by the shaded bands clearly reflect their expected behavior, insofar as the error bands are narrowly focused about their mean (but not unreasonably so) where AME data exists, whereas the predictions begin to drift past the limits of the training and testing dataset, the same uncertainty bands begin to diverge, up to about $3\sigma \approx 1$ MeV for the isotopes shown. This opens up the possibility for more thorough uncertainty quantification analyses based on this general approach, which we intend to explore in future works.

An important consequence of PIML modeling is that the output of the network may be interpreted and understood as with any useful theoretical model. To this end, we compute Shapley additive explanations (SHAP) values [51] to measure feature importance. Figure 3 ranks the eight features when applied to the more recent AME2020 [44]. We find that macroscopic quantities rank the highest in determining the masses, followed by the features which control the quantum effects. Figure 3 reflects the long held belief that, to first order, the atomic nucleus is well described by bulk macroscopic features, whereas microscopic features induce subtle, yet extremely important corrections [52,53]. This is evident when comparing a nucleus with its close neighbors where the macroscopic features may be very nearly equal, but the quantum effects in these nuclei can lead to dramatic differences in nearly every nuclear observable.

Conclusions. We present a probabilistic machine learning algorithm with the capacity to directly predict the nuclear binding of atomic nuclei. We achieve an unprecedented match to training data with a root-mean-square error of 186 keV ($\approx 20\%$ of the AME) whereas utilizing a single physical constraint and only eight parameters defining the feature space.
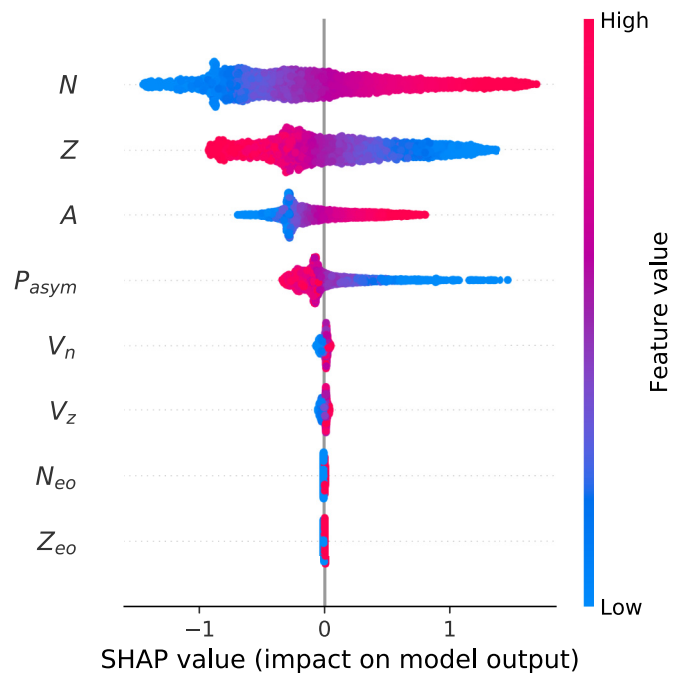

FIG. 3. The ranking of feature importance measured by SHAP value over the entire AME2020 [44]. Macroscopiclike terms rank the highest followed by the quantum effects of Pauli exclusion, valence nucleons, and pairing.

Our model is capable of predicting all of the AME2016 with $Z \geqslant 20$ at $\sigma_{\text{rms}} \approx 316$ keV.

Our PIML approach affords the ability to analyze manifestations of physical phenomena. We demonstrate this by establishing relative importance of macroscopic and microscopic input features which is consistent with the traditional understanding of the atomic nuclear binding energy viz. the semiempirical mass formula as well as with macroscopic-microscopic and fully microscopic methods. Figure 3 reinforces this viewpoint by ranking the eight features implemented in this Letter using the game theoretic SHAP measure [54]. Importantly, it is found that macroscopic features are ranked as most influential since they determine the bulk of nuclear binding, followed by features that are more closely associated with quantum corrections. This calculation is performed over the entire predictive range of the AME2020 nuclei and shown to retain this behavior. Study of this behavior far from measured isotopes will help to assess the extrapolation quality of this procedure and is the subject of future work.

The technology developed here is general and may, in principle, be applied to the study of any physical observable or combination thereof. A concerted effort in this approach opens new possibilities to capture complex physics that is not currently viable via other contemporary approaches to computational physics. This may prove especially valuable in advancing the study of a wide range of many-body problems that permeate physics.

Data and Software. The ML mass model predictions associated with this Letter are available upon request to the corresponding author. The software (source code) used

[1] M. Thoennessen, Plans for the facility for rare isotope beams, Nucl. Phys. A **834**, 688c (2010).

[2] P. Talou, I. Stetcu, and T. Kawano, Modeling the emission of prompt fission gamma-rays for fundamental physics and applications, Phys. Procedia **59**, 83 (2014).

[3] F. Kondev, M. Wang, W. Huang, S. Naimi, and G. Audi, The NUBASE2020 evaluation of nuclear physics properties, Chin. Phys. C **45**, 030001 (2021).

[4] T. Kawano, R. Capote, S. Hilaire, and P. Chau Huu-Tai, Statistical hauser-feshbach theory with width-fluctuation correction including direct reaction channels for neutron-induced reactions at low energies, Phys. Rev. C **94**, 014612 (2016).

[5] J. Randrup and R. Vogt, Generation of Fragment Angular Momentum in Fission, Phys. Rev. Lett. **127**, 062502 (2021).

[6] A. W. Steiner, Deep crustal heating in a multicomponent accreted neutron star crust, Phys. Rev. C **85**, 055804 (2012).

[7] G. W. Misch, B. A. Brown, and G. M. Fuller, Neutrino-pair emission from hot nuclei during stellar collapse, Phys. Rev. C **88**, 015807 (2013).

[8] M. Mumpower, R. Surman, G. McLaughlin, and A. Aprahamian, The impact of individual nuclear properties on r-process nucleosynthesis, Prog. Part. Nucl. Phys. **86**, 86 (2016).

[9] T. M. Sprouse, R. Navarro Perez, R. Surman, M. R. Mumpower, G. C. McLaughlin, and N. Schunck, Propagation of statistical uncertainties of skyrme mass models to simulations of r-process nucleosynthesis, Phys. Rev. C **101**, 055803 (2020).

[10] Y. L. Zhu, K. A. Lund, J. Barnes, T. M. Sprouse, N. Vassh, G. C. McLaughlin, M. R. Mumpower, and R. Surman, Modeling Kilonova Light Curves: Dependence on Nuclear Inputs, Astrophys. J. **906**, 94 (2021).

[11] P. Möller, A. J. Sierk, T. Ichikawa, and H. Sagawa, Nuclear ground-state masses and deformations: FRDM(2012), At. Data Nucl. Data Tables **109-110**, 1 (2016).

[12] S. Goriely, S. Hilaire, M. Girod, and S. Péru, First Gogny-Hartree-Fock-Bogoliubov Nuclear Mass Model, Phys. Rev. Lett. **102**, 242501 (2009).

[13] R. Utama, J. Piekarewicz, and H. B. Prosper, Nuclear mass predictions for the crustal composition of neutron stars: A bayesian neural network approach, Phys. Rev. C **93**, 014311 (2016).

[14] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, Bayesian approach to model-based extrapolation of nuclear observables, Phys. Rev. C **98**, 034318 (2018).

[15] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, Hyperparameter optimization for machine learning models based on bayesian optimizationb, J. Electron. Sci. Technol. **17**, 100007 (2019).

[16] A. Boehnlein, M. Diefenthaler, C. Fanelli, M. Hjorth-Jensen, T. Horn, M. P. Kuchera, D. Lee, W. Nazarewicz, K. Orginos, P. Ostroumov, L.-G. Pang, A. Poon, N. Sato, M. Schram, A. Scheinker, M. S. Smith, X.-N. Wang, and V. Ziegler, Machine Learning in Nuclear Physics, arXiv:2112.02309 (2021).

[17] G. A. Negoita, J. P. Vary, G. R. Luecke, P. Maris, A. M. Shirokov, I. J. Shin, Y. Kim, E. G. Ng, C. Yang, M. Lockner, and G. M. Prabhu, Deep learning: Extrapolation tool for *ab initio* nuclear theory, Phys. Rev. C **99**, 054308 (2019).

[18] W. G. Jiang, G. Hagen, and T. Papenbrock, Extrapolation of nuclear structure observables with artificial neural networks, Phys. Rev. C **100**, 054326 (2019).

[19] K. Raghavan, P. Balaprakash, A. Lovato, N. Rocco, and S. M. Wild, Machine-learning-based inversion of nuclear responses, Phys. Rev. C **103**, 035502 (2021).

[20] N. Ismail and A. Gezerlis, Machine-learning approach to finite-size effects in systems with strongly interacting fermions, Phys. Rev. C **104**, 055802 (2021).

[21] V. Sobes, M. Grosskopf, K. A. Wendt, D. Brown, M. Smith, and P. Talou, WANDA: AI/ML for Nuclear Data, Oak Ridge National Lab. (ORNL), Oak Ridge,Tech. Rep. No. ORNL/TM-2020/1535 (2020).

[22] K. Kolos, V. Sobes, R. Vogt, C. E. Romano, M. S. Smith, L. A. Bernstein, D. A. Brown, M. T. Burkey, Y. Danon, M. A. Elsawi, B. L. Goldblum, L. H. Heilbronn, S. L. Hogle, J. Hutchinson, B. Loer, E. A. McCutchan, M. R. Mumpower, E. M. O'Brien, C. Percher, P. N. Peplowski *et al.*, Current nuclear data needs for applications, Phys. Rev. Research **4**, 021001 (2022).

[23] L. Neufcourt, Y. Cao, W. Nazarewicz, E. Olsen, F. Viens, Neutron Drip Line in the Ca Region from Bayesian Model Averaging, Phys. Rev. Lett. **122**, 062502 (2019).

[24] L. Neufcourt, Y. Cao, S. A. Giuliani, W. Nazarewicz, E. Olsen, and O. B. Tarasov, Quantified limits of the nuclear landscape, Phys. Rev. C **101**, 044307 (2020).

[25] M. Shelley and A. Pastore, A new mass model for nuclear astrophysics: Crossing 200 kev accuracy, Universe **7**, 131 (2021).

[26] A. Sharma, A. Gandhi, and A. Kumar, Learning correlations in nuclear masses using neural networks, Phys. Rev. C **105**, L031306 (2022).

[27] A. Pastore and M. Carnini, Extrapolating from neural network models: a cautionary tale, J. Phys. G: Nucl. Part. Phys. **48**, 084001 (2021).

[28] Z. Niu and H. Liang, Nuclear mass predictions based on bayesian neural network approach with pairing and shell effects, Phys. Lett. B **778**, 48 (2018).

[29] U. Murarka, K. Banerjee, T. Malik, and C. Providência, The neutron star outer crust equation of state: a machine learning approach, J. Cosmol. Astropart. Phys. **2022** 045 (2022).

[30] H. Özdoğan, Y. A. Üncü, M. Şekerci, and A. Kaplan, Mass excess estimations using artificial neural networks, Appl. Radiat. Isot. **184**, 110162 (2022).

[31] L. A. Bernstein, D. A. Brown, A. J. Koning, B. T. Rearden, C. E. Romano, A. A. Sonzogni, A. S. Voyles, and W. Younes, Our future nuclear data needs, Annu. Rev. Nucl. Part. Sci. **69**, 109 (2019).

[32] A. E. Lovell, A. T. Mohan, T. M. Sprouse, and M. R. Mumpower, Nuclear masses learned from a probabilistic neural network, Phys. Rev. C **106**, 014305 (2022).

[33] C. M. Bishop, Mixture Density Networks, Technical Report, Aston University 1994).

[34] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai *et al.*, Pytorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alche Buc, E. Fox, and R. Garnett (Curran, Red Hook, NY, USA, 2019), pp. 8024–8035.

[35] A. E. Lovell, A. T. Mohan, and P. Talou, Quantifying uncertainties on fission fragment mass yields with mixture density networks, J. Phys. G: Nucl. Part. Phys. **47**, 114001 (2020).

[36] R. Casten and N. Zamfir, The evolution of nuclear structure: The npnn scheme and related correlations, J. Phys. G: Nucl. Part. Phys. **22**, 1521 (1996).

[37] R. F. Casten, D. S. Brenner, and P. E. Haustein, Valence *p*-*n* Interactions and the Development of Collectivity in Heavy Nuclei, Phys. Rev. Lett. **58**, 658 (1987).

[38] M. Wang, G. Audi, F. G. Kondev, W. Huang, S. Naimi, and X. Xu, The AME2016 atomic mass evaluation (II). tables, graphs and references, Chin. Phys. C **41**, 030003 (2017).

[39] G. T. Garvey, W. J. Gerace, R. L. Jaffe, I. Talmi, and I. Kelson, Set of nuclear-mass relations and a resultant mass table, Rev. Mod. Phys. **41**, S1 (1969).

[40] J. Barea, A. Frank, J. G. Hirsch, P. V. Isacker, S. Pittel, and V. Velázquez, Garvey-kelson relations and the new nuclear mass tables, Phys. Rev. C **77**, 041304(R) (2008).

[41] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[42] K. Deb, Multi-objective optimization, in *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*, edited by E. K. Burke and G. Kendall (Springer, Boston, 2005), pp. 273–316.

[43] N. Gunantara, A review of multi-objective optimization: Methods and its applications, Cogent Eng. **5**, 1502242 (2018).

[44] M. Wang, W. Huang, F. Kondev, G. Audi, and S. Naimi, The AME 2020 atomic mass evaluation (II). tables, graphs and references, Chin. Phys. C **45**, 030003 (2021).

[45] R. Orford, N. Vassh, J. A. Clark, G. C. McLaughlin, M. R. Mumpower, G. Savard, R. Surman, A. Aprahamian, F. Buchinger, M. T. Burkey, D. A. Gorelov, T. Y. Hirsh, J. W. Klimes, G. E. Morgan, A. Nystrom, and K. S. Sharma, Precision Mass Measurements of Neutron-Rich Neodymium and Samarium Isotopes and Their Role in Understanding Rare-Earth Peak Formation, Phys. Rev. Lett. **120**, 262702 (2018).

[46] J. Jänecke and P. Masson, Mass predictions from the garvey-kelson mass relations, At. Data Nucl. Data Tables **39**, 265 (1988).

[47] D. Lunney, J. M. Pearson, and C. Thibault, Recent trends in the determination of nuclear masses, Rev. Mod. Phys. **75**, 1021 (2003).

[48] I. O. Morales, J. C. López Vieyra, J. G. Hirsch, and A. Frank, How good are the Garvey-Kelson predictions of nuclear masses?, Nucl. Phys. A **828**, 113 (2009).

[49] J. A. Melendez, R. J. Furnstahl, D. R. Phillips, M. T. Pratola, and S. Wesolowski, Quantifying correlated truncation errors in effective field theory, Phys. Rev. C **100**, 044001 (2019).

[50] N. Schunck, J. ONeal, M. Grosskopf, E. Lawrence, and S. M. Wild, Calibration of energy density functionals with deformed nuclei, J. Phys. G: Nucl. Part. Phys. **47**, 074001 (2020).

[51] S. Lundberg and S.-I. Lee, A unified approach to interpreting model predictions, arxiv:1705.07874.

[52] C. F. V. Weizsäcker, Zur Theorie der Kernmassen, Z. Phys. **96**, 431 (1935).

[53] J. W. Negele, The mean-field theory of nuclear structure and dynamics, Rev. Mod. Phys. **54**, 913 (1982).

[54] L. S. Shapley, 17. a value for n-person games, in *Contributions to the Theory of Games (AM-28)*, edited by H. W. Kuhn and A. W. Tucker (Princeton University Press, Princeton, NJ, USA, 2016), Vol. II, pp. 307–318.