

Physics-guided ensemble learning of nuclear masses for astrophysical applications

Srikrishna Agrawal ^{*}

Department of Mechanical Engineering, BITS-Pilani Hyderabad Campus, Hyderabad 500078, India

N. Chandnani 

*Department of Physics, Manipal University Jaipur, Jaipur 303007, India
and Department of Physics (H&S), Government Women Engineering College, Ajmer 305002, India*

T. Ghosh  and N. Paar 

Department of Physics, Faculty of Science, University of Zagreb, Bijenička cesta 32, Zagreb 10000, Croatia

G. Saxena [†]

Department of Physics (H&S), Government Women Engineering College, Ajmer 305002, India

B. K. Agrawal 

Saha Institute of Nuclear Physics, 1/AF, Bidhannagar, Kolkata 700064, India



(Received 10 April 2026; accepted 14 July 2026; published 6 August 2026)

The gradient boosting regressor, an ensemble learning technique, is utilized to correct the residuals in nuclear mass excess across six different nuclear mass models. The corrected residuals are substantially reduced and further improved by applying a weighted averaging method, which benefits from partial error cancellation among the models. Our proposed conflated model, combining ensemble learning with model averaging (ELMA), achieves a root mean square error (RMSE) of approximately 57 keV, well below the critical threshold of 100 keV, for the full dataset of the Atomic Mass Evaluation (AME2020). ELMA also improves predictions for nuclei beyond the training region, with the weighted ensemble reducing the RMSE by approximately 25%–45% relative to the best individual mass model on independent extrapolative datasets. The reliability of the ELMA approach is further validated through the calculation of α -decay Q values, where a clear reduction in deviations from experimental data is observed compared to individual nuclear mass model predictions. We have also compiled a mass excess and binding energy of 6300 nuclei, obtained using our framework and serves as a valuable resource for nuclear physics research and is publicly available through the ELMA web interface (<https://ddnp.in/>).

DOI: [10.1103/8n1w-f3b7](https://doi.org/10.1103/8n1w-f3b7)

^{*}Contact author: agrawalsrikrishna0411@gmail.com

[†]Contact author: gauravphy@gmail.com

I. INTRODUCTION

The mass of an atomic nucleus, or its mass excess, is one of the fundamental observables that encapsulates the total binding energy, thereby serving as a direct insight into the underlying nuclear interactions and many-body quantum structure [1]. Accurate values of mass are necessary for mapping the limits of nuclear existence at the proton and neutron drip lines [1]. It also plays an important role in astrophysical reaction network calculations, determining nuclear abundances not only through rapid and slow neutron capture processes but also in other nucleosynthesis pathways. The uncertainties in nuclear masses of a few hundred keV can change calculated elemental abundances by many orders of magnitude [1,2].

Despite several technological advances in mass measurements [3–7], experimental determination is exceedingly challenging for nuclei far from the line of stability, which are characterized by extremely short half-lives, low production rates [1], and the complicated task of separating low-energy isomeric states from the ground state [8]. Therefore, one relies heavily on theoretical calculations for the unexplored territory of the nuclear chart [9].

The Atomic Mass Evaluation (AME2020) provides a comprehensive database of experimentally measured masses for over 2400 nuclei [10]; however, theoretical predictions remain indispensable for the over 6000 predicted isotopes [11,12] that are experimentally inaccessible. Since the 1930s, a multitude of models spanning macroscopic-microscopic and fully microscopic frameworks [13–24] have progressively reduced the root mean square error (RMSE) from 3 MeV for the earliest liquid drop model to 0.2–0.8 MeV for the most recent global models. Despite this progress, their predictive performance deteriorates for the nuclei far from the β -stability line, where the deviations can increase to several MeV [25]. This uncertainty poses a major challenge to consistent extrapolation of exotic nuclear properties and far exceeds the sub-100 keV precision needed to perform high-fidelity r -process calculations [26].

This accuracy gap has motivated the application of machine learning (ML) via residual learning, where algorithms are trained on the difference between experimental and theoretical mass values [9]. Pioneered by Utama *et al.* [27] using a Bayesian neural network (BNN), this strategy has since been extended through a diverse range of architectures, including Gaussian process regression (GPR), physics-guided neural networks, advanced BNN models, and tree-based ensemble methods [28–32], with the RMSE progressively reducing from 0.374 to 0.191 MeV on the validation set and from 0.260 to 0.090 MeV on the AME2020 dataset. However, reliance on a single best model introduces model uncertainty and overconfident inferences [33–35], a challenge underscored by the No Free Lunch theorem [36]. Model averaging through ensemble methods addresses this directly: By aggregating diverse models, each capturing different facets of the data, the ensemble reduces overall error, mitigates individual model bias, and improves generalization [34,35,37]. This is illustrated by Bentley *et al.* [38], whose four model tree ensemble (FMTE), formed by a weighted average of four distinct boosted tree models, achieved an RMSE of 0.076 MeV against AME2020 and 0.376 MeV for post-AME2020 extrapolations, lower than the individual mass models considered.

Unlike previous residual learning studies that focus on a single nuclear mass model, the present work combines machine-learned residual corrections from six theoretically distinct global mass models within a unified ensemble framework, with particular emphasis on predictive performance in extrapolative regions of the nuclear chart. We employ a diverse set of six nuclear mass models broadly characterized as macroscopic-microscopic, nonrelativistic, and relativistic self-consistent mean-field models. The RMSE for the residuals of nuclear masses for these models over the full AME2020 is in the range 0.25–1.0 MeV. We have corrected raw residuals (RRs) obtained from individual nuclear mass models using the gradient boosting regressor (GBR) [39], one of the prominent ensemble learning methods [31,40]. GBR builds models sequentially, where each model is trained to correct the errors of its predecessors, resulting in reduced bias [41]. We obtain a single improved nuclear mass model by performing a weighted average of all corrected residuals (CRs). Our conflated nuclear mass model obtained by integrating ensemble learning with model averaging (ELMA) yields an overall RMSE 57 keV, well below the critical value of 100 keV. We also confront

our mass model with newly measured mass excesses and with the Q values of α -decay. We have compiled values of nuclear masses and binding energies for about 6000 nuclei from our improved nuclear mass model.

II. FEATURES AND DATASET

A robust foundation for any machine learning model lies in both the selection of relevant input features and the careful partitioning of the dataset for training, testing, and extrapolative evaluation. We employ a set of 12 physics-guided features, chosen to capture essential characteristics of nuclear structure and empirical regularities known to influence nuclear mass excess. These features were chosen based on their relevance to nuclear structure and their proven efficacy in previous mass modeling studies [28]. The fundamental features include the proton number (Z) and neutron number (N), which specify the nuclide and incorporate well-established shell closures intrinsic to nuclear stability. The mass number (A) and its transformations, such as $A^{2/3}$, encode macroscopic size and surface effects consistent with liquid drop-inspired mass formulas.

The isospin asymmetry [$I = (N - Z)/A$] measures the difference between the number of neutrons and protons in a nucleus, and it plays an important role in the symmetry energy part of the nuclear equation of state. To account for pairing effects, which cause sudden changes in mass between nuclei with odd or even numbers of protons and neutrons (known as even-odd mass staggering), we include binary variables (Z_{eo} , N_{eo}) indicating whether the proton and neutron numbers are even or odd.

The magic number gap features (v_Z , v_N) characterize shell effects and are central to explaining localized maxima and minima of nuclear mass surfaces, quantifying how close a nucleus is to canonical magic numbers such as 8, 20, 28, 50, 82, 126, and 184. These features enable the model to implicitly identify shell closures and the associated enhanced stability. In addition, the promiscuity factor (PF), defined as $PF = \frac{v_Z v_N}{v_Z + v_N}$, captures the interplay between valence protons and neutrons and serves as an indicator for proton-neutron interaction strength. This feature helps to describe collective effects beyond simple shell closures. Finally, the proton and neutron shell numbers (Z_{shell} , N_{shell}) provide discrete indices representing the shell model orbitals of the last proton and neutron. These indices (0–4) map the nucleon numbers into established shell ranges, such as 1–28, 29–50, 51–82, 83–126, and above 127, thereby encoding large-scale structural organization of the nuclear chart. Such a comprehensive set of physics-guided features enables the models to learn complex patterns, including shell closures and structural irregularities that are often missed by purely empirical predictors.

The complete dataset used in this work consists of 2488 nuclei. Of these, 2386 experimentally known nuclei from AME2020 were used for the GBR residual-learning procedure. They were randomly divided into training (85%) and test (15%) sets, as shown in Fig. 1. The training set was used to train the GBR model, while the test set was used to evaluate its performance on known nuclei. The remaining 102 nuclei were excluded from model development. Among these, the 71 nuclei whose masses were newly incorporated in AME2020 (extpl) were kept completely separate from the training and test sets and used only to assess the model’s extrapolative performance. To further evaluate its predictive capability, following Ref. [38], we also include 31 mass measurements reported after AME2020 (new) as a second independent test set. The datasets employed for training and validation are summarized in Table I.

This setup allows us to fairly compare how well the models work on both well-known nuclei and on new, less explored ones.

III. GRADIENT BOOSTING REGRESSOR

The gradient boosting regressor builds a strong predictive ensemble by fitting a series of regressors sequentially to the residuals of earlier models rather than to the target values directly. At each stage, a new regressor is trained to predict the negative gradient of the loss function, iteratively

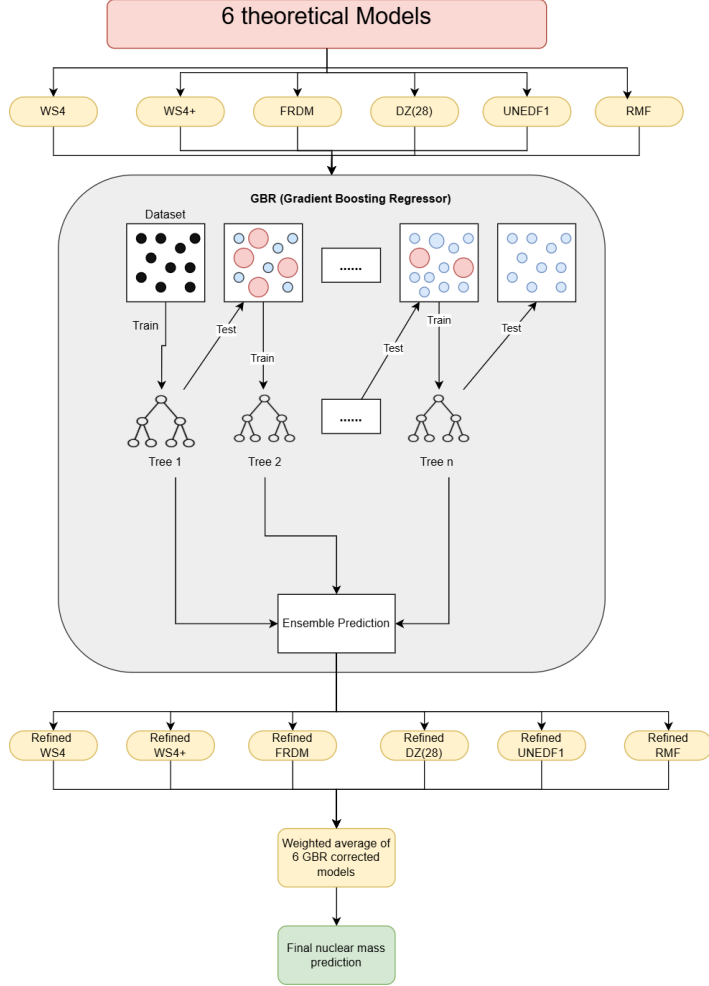


FIG. 1. The raw residuals of six nuclear mass models, WS4, WS4+, FRDM, DZ(28), UNEDF1, and RMF, are each corrected independently using GBR, which builds a sequence of 3000 sequential decision-tree estimators trained on the residual errors of their predecessors. The six sets of refined residuals are then combined via a normalized weighted average to yield the final nuclear mass prediction.

minimizing it in a stagewise manner, thereby capturing the complex nonlinear dependencies present in nuclear mass excess data [39]. The contribution of each new estimator is scaled by a learning rate, which controls the magnitude of each incremental update, a smaller learning rate generally leads to

TABLE I. Summary of datasets used for training, validation, and extrapolative evaluation of the model. The table lists the number of nuclei in each subset: training set (train), test set (test), extrapolated nuclei from AME2020 (extpl), the complete AME2020 dataset (2020), nuclei reported for the first time after the compilation of AME2020 (new), and the full dataset comprising all nuclei considered.

Set	Train	Test	Extpl	2020	New	Full
Number	2028	358	71	2457	31	2488

better generalization at the cost of requiring more iterations. A key distinguishing feature of GBR is its flexibility in optimizing arbitrary differentiable loss functions, with mean squared error serving as the default for regression tasks.

In the present work, we used the GBR implementation from the scikit-learn library with `n_estimators = 3000`, `learning_rate = 0.1`, `max_depth = 3`, and `subsample = 0.8`. The value of `n_estimators` is determined through a convergence study in which the weighted-average validation RMSE monitored as a function of boosting stages; all validation metrics plateau beyond 3000 stages with changes of less than 0.001 MeV, confirming convergence while preserving optimal generalization capability. The remaining hyperparameters follow standard practice for GBR regression on structured tabular data [39,40,42,43].

The complete workflow of the framework is illustrated in Fig. 1. The raw residuals of each of the six nuclear mass models are independently passed through the GBR pipeline, which builds a sequence of 3000 decision-tree estimators, each trained to correct the residual errors of its predecessors. The ensemble prediction for each model is obtained by combining all estimators weighted by the learning rate, yielding six sets of refined residuals. These are then combined through a weighted average to produce the final nuclear mass prediction. The computational cost of the framework scales approximately linearly with the number of nuclear mass models, while inference remains computationally inexpensive.

IV. RESULTS AND DISCUSSION

We have selected six nuclear mass models, namely, Weizsäcker-Skyrme (WS4) [17], WS4+ [18], finite-range droplet model (FRDM) [15], Duflo Zuker model (DZ28) [16], Universal Nuclear Energy Density Functional 1 (UNEDF1) [44], and relativistic mean field (RMF) [45,46] that represent a wide spectrum of theoretical approaches. The WS4 and its improved version WS4+ are macroscopic-microscopic models based on the Skyrme energy-density functional, with WS4+ including further empirical corrections for enhanced accuracy. The FRDM combines macroscopic finite range droplet energy and microscopic shell and pairing corrections, providing reliable predictions for masses across the nuclear chart. The DZ(28) is a parameter-rich Duflo-Zuker model that blends macroscopic and microscopic features for improved mass reproduction. The UNEDF1 is a self-consistent nonrelativistic mean-field model based on an effective Skyrme energy-density functional for global modeling of nuclear properties. The RMF model describes the nucleus by treating nucleons as interacting through the exchange of mesons under the mean-field approximation, with nonlinear self-interaction terms included in the meson fields. These models collectively cover both empirical and fully microscopic theoretical frameworks, providing a broad and robust basis for our present analysis.

To study the baseline predictive performance of the selected nuclear mass models, we computed the RR in nuclear mass excess as the difference between measured and theoretical values for each model across six different datasets, as listed in Table I. This observation is consistent with the visual residual distribution in Fig. 2, where we have displayed RR for all the nuclei considered in various datasets across the nuclear chart. These findings suggest that nuclei in the well-characterized regions near the β -stability line (squares) appear predominantly in green, indicating residuals close to zero. As one moves away from the stability line, the color distribution becomes progressively more mixed, shifting toward blue and red. This trend is most pronounced for the extrapolated (diamonds) and newly measured (triangles) nuclei, which show the largest deviations and the least agreement across models.

To improve the accuracy of nuclear mass prediction, the RR are corrected using an ensemble learning approach, GBR. The RR is defined as

$$\delta M_i^{\text{RR}} = M_{\text{Exp}} - M_i, \quad (1)$$

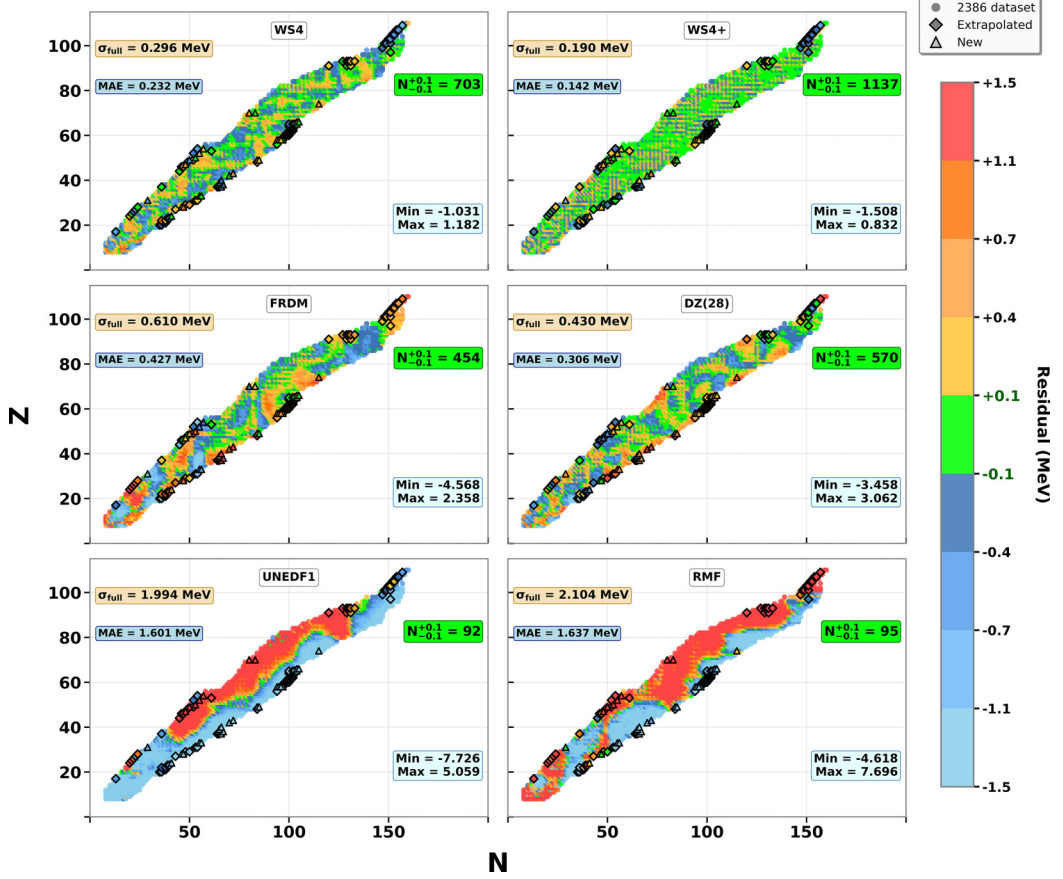


FIG. 2. The raw residuals in nuclear mass excess (ΔM_i^{RR}), defined as the deviation between measured and theoretical values, are presented as functions of neutron number (N) and proton number (Z) for six nuclear mass models: WS4, WS4+, FRDM, DZ(28), UNEDF1, and RMF. The panels display the residuals for 2386 previously known nuclei (circles) and residuals for an extrapolated set of 71 newly measured nuclei added only in AME2020 (diamonds). Thirty-one newly measured nuclei after AME2020 are shown by triangles. A blue-to-red color scale represents the residuals: blue indicates overbinding (theoretical mass lower than experimental), red indicates underbinding, and green-yellow corresponds to values near zero. Each subplot includes the values of standard deviation and the number of nuclei lying in the green region $-0.1 \leq \Delta M_i^{\text{RR}} \leq +0.1$ MeV.

where M_{Exp} is the experimental nuclear mass excess for a given nucleus and M_i is the corresponding value for the i th nuclear mass model. The CR are

$$\delta M_i^{\text{CR}} = \delta M_i^{\text{RR}} - \delta M_i^{\text{ML}} \quad (2)$$

$$= M_{\text{Exp}} - M_i^{\text{cor}}, \quad (3)$$

where the corrected mass excess $M_i^{\text{cor}} = M_i + \delta M_i^{\text{ML}}$, with δM_i^{ML} being the machine-learned value of the raw residuals. We employ GBR to estimate δM_i^{ML} in Eq. (2). The weighted average of residuals is obtained as

$$\delta \bar{M}^{\text{X}} = \sum_{i=1}^{N_{\text{X}}} w_i^{\text{X}} \delta M_i^{\text{X}}, \quad (4)$$

where X denotes either RR or CR, with $N_{\text{RR/CR}} = N_m$, and N_m the number of nuclear mass models considered. The normalized weights w_i^X are given by

$$w_i^X = \frac{\left(\frac{1}{\sigma_i^X}\right)^2}{\sum_{k=1}^{N_m} \left(\frac{1}{\sigma_k^X}\right)^2}, \quad (5)$$

where σ_i^X is the RMSE for a given nuclear mass model with raw or corrected residuals, evaluated as

$$\sigma_i^X = \sqrt{\frac{1}{N_d} \sum_{j=1}^{N_d} (\delta M_i^X(j))^2}, \quad (6)$$

where the index j runs over all nuclei of a given dataset of size N_d . Thus, models with smaller values of σ_i^X receive higher weights, contributing more effectively to improving the accuracy of the prediction. These weights are calculated using the RMSE values from the test dataset.

The models were trained using 12 physics-guided features, as detailed in Sec. II, with the target variable being δM_i^{RR} as given by Eq. (1). The training dataset is constructed from the Atomic Mass Evaluation (AME2020) by excluding the nuclear mass excess values of 71 nuclei measured for the first time, as described in Sec. II. After training, the model performance for the corrected residuals is evaluated on various datasets—labeled as test, extpl, 2020, new, and full.

The robustness of the results with respect to the train-test data split has been examined using several commonly adopted partitions, as reported in Sec. IV. Based on this analysis, an 85/15 train-test split has been adopted for all subsequent calculations.

The RMSE values (σ) obtained using this chosen split for all six nuclear mass models, along with their normalized weights and weighted averages evaluated across all datasets, are listed in Table II. For the RR, the weight distribution is heavily skewed: WS4+ alone carries approximately 62% of the total weight and together with WS4 accounts for $\sim 82\%$, leaving only $\sim 18\%$ distributed among the remaining four models: FRDM, DZ(28), UNEDF1, and RMF. This strong dominance directly reflects the significantly lower RMSE of WS4 and WS4+ relative to the other models.

After GBR correction, however, the weight distribution becomes substantially more equitable. For the CR, WS4+ contributes $\sim 34\%$ and WS4 contributes $\sim 20\%$, so that together they account for only $\sim 54\%$ of the total weight, while the remaining four models collectively contribute $\sim 46\%$. This redistribution arises because the GBR algorithm, by learning and correcting the residual patterns specific to each model, suppresses the large model-dependent biases that drove the disparity in RMSE values. As a result, the corrected residuals of the remaining models, such as FRDM, DZ(28), UNEDF1, and RMF, are brought considerably closer to those of WS4 and WS4+, so that the weighted average via Eq. (4) draws more meaningfully from the full diversity of theoretical approaches rather than being dominated by a single model. This broader participation is particularly valuable in regions where no single model performs reliably and by spreading the weight across multiple models, the ensemble reduces the risk of large errors that would arise from over-reliance on any one of them. To assess the effects of model correlations, we also investigated reduced ensembles obtained by excluding correlated mass models. Although the performance for the known nuclei remains comparable, the complete six-model ensemble provides slightly more consistent predictions for nuclei outside the training region, suggesting that each model contributes complementary predictive information.

The single best-performing model for RR, WS4+, achieves a test RMSE of 0.1735 MeV, which is reduced to 0.1208 MeV for CR, corresponding to an improvement of $\sim 30\%$. The weighted average over all six models through Eq. (4) yields a test RMSE of 0.1976 MeV for RR and 0.1245 MeV for CR, corresponding to an improvement of $\sim 37\%$. Although the weighted CR average (0.1245 MeV) is only marginally higher than the best single-model CR value (0.1208 MeV) for the test dataset, its benefit becomes evident for nuclei beyond the training region. For the extpl dataset, the

TABLE II. Comparison of root-mean-square errors (σ , in MeV) of raw mass-excess residuals (RR) and corrected residuals (CR) for the mass excess of six nuclear mass models, evaluated for the training, test, extrapolated (extpl), AME2020, newly measured (new), and full datasets as described in Table I. The last column lists the corresponding weights (W) derived from the test-set RMSE. The rows in italic list the weighted average of the six nuclear mass models.

	σ (MeV)						
Models	Train	Test	Extpl	2020	New	Full	W
Raw residuals (RR)							
WS4	0.2895	0.3092	0.3556	0.2946	0.3713	0.2956	0.1954
WS4+	0.1840	0.1735	0.3401	0.1889	0.2893	0.1905	0.6204
FRDM	0.6040	0.5755	0.7864	0.6060	0.8478	0.6096	0.0564
DZ(28)	0.4238	0.3972	0.6330	0.4276	0.5782	0.4298	0.1184
UNEDF1	1.9962	1.9169	1.9526	1.9836	2.6935	1.9940	0.0051
RMF	2.0756	2.0972	2.8321	2.1044	2.0882	2.1042	0.0042
<i>Weighted average</i>	<i>0.2004</i>	<i>0.1976</i>	<i>0.3285</i>	<i>0.2049</i>	<i>0.2815</i>	<i>0.2060</i>	
Corrected residuals (CR)							
WS4	0.0228	0.1593	0.2214	0.0745	0.2467	0.0790	0.1987
WS4+	0.0187	0.1208	0.2104	0.0608	0.2667	0.0673	0.3458
FRDM	0.0277	0.1726	0.2855	0.0856	0.2978	0.0914	0.1693
DZ(28)	0.0245	0.1760	0.2697	0.0843	0.2946	0.0900	0.1629
UNEDF1	0.0304	0.2763	0.4379	0.1320	0.4040	0.1387	0.0661
RMF	0.0480	0.2972	0.5711	0.1555	0.5210	0.1651	0.0571
<i>Weighted average</i>	<i>0.0179</i>	<i>0.1245</i>	<i>0.1816</i>	<i>0.0590</i>	<i>0.2141</i>	<i>0.0633</i>	

weighted CR average is $\sim 14\%$ lower than the best single-model CR value, while for the new dataset the corresponding reduction is $\sim 20\%$. Across all datasets, the weighted CR (σ) values remain lower than the corresponding RR (σ) values. The reduction ranges from $\sim 96\%$ for the training dataset and $\sim 37\%$ for the test dataset to $\sim 72\%$ and $\sim 70\%$ for the 2020 and full datasets, respectively. For the extpl and new datasets, the corresponding reductions are $\sim 45\%$ and $\sim 24\%$, respectively.

Figure 3 shows the probability density distributions of the weighted-average raw and corrected residuals for the test, extpl, and new set of nuclei. For the test dataset, both $\overline{RR}$ and $\overline{CR}$ are centered close to zero, but the weighted average distribution for the corrected residual is sharper, with a peak density nearly twice that of the weighted average distribution of raw residuals, indicating a reduction in residual spread after GBR correction. A similar behavior is observed for the extpl dataset, where the CR distribution is narrower and more symmetric about zero compared to the broader RR distribution, consistent with the $\sim 45\%$ reduction in σ reported in Table II. For the new dataset, the GBR correction continues to improve predictive stability, with the CR distribution remaining sharper despite the increased complexity of nuclei far from the β -stability line. The mild asymmetry with a tail toward positive residuals is a natural consequence of this complexity, while the $\sim 24\%$ reduction in σ confirms that the model retains meaningful corrective capability even in this extrapolation region. The extrapolative performance of GBR compares favourably with existing machine-learning-based nuclear mass study. For the 71-nuclei extrapolated set, Yüksel *et al.* [28] report 0.670 MeV using GPR, while ELMA achieves 0.182 MeV, a reduction of $\sim 73\%$.

The spatial distribution of these improvements across the nuclear chart is displayed in Fig. 4, which shows the weighted-average raw residuals $\delta\bar{M}^{\text{RR}}$ (lower panel) and corrected residuals $\delta\bar{M}^{\text{CR}}$ (upper panel) for all nuclei in the complete dataset. Nuclei whose absolute residuals fall within 100 keV are shown in green, with their total count denoted by $N_{-0.1}^{+0.1}$. An increase in $N_{-0.1}^{+0.1}$ is observed after GBR correction, with the majority of nuclei across all three datasets, including those

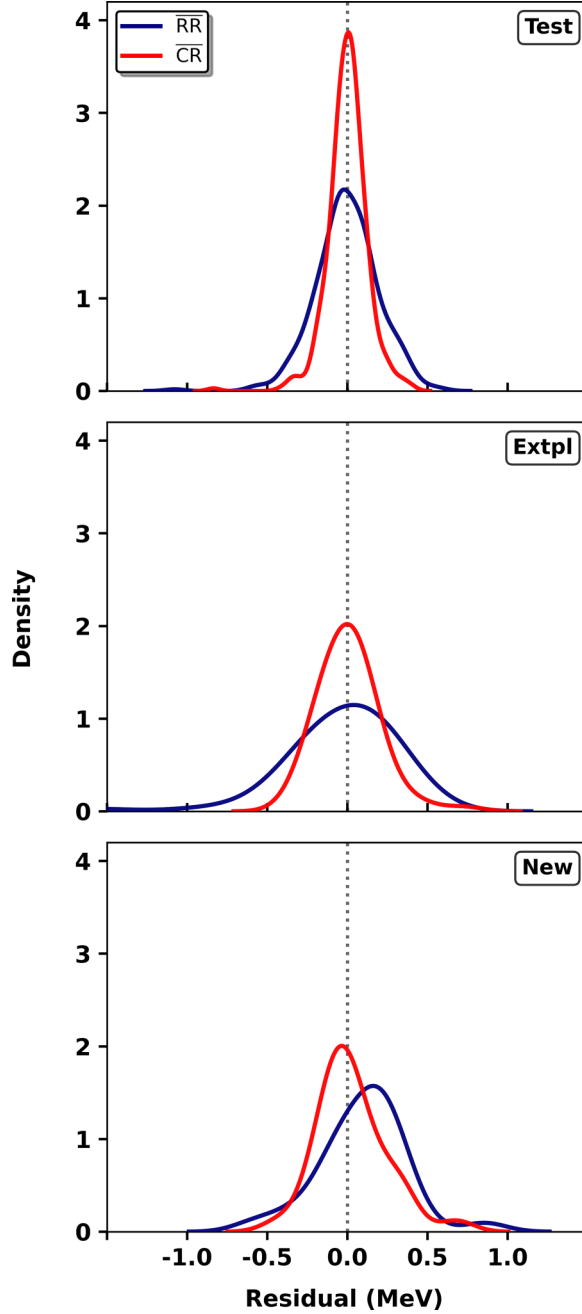


FIG. 3. Probability density distributions of the weighted-average raw residuals ($\overline{RR}$, blue) and corrected residuals ($\overline{CR}$, red) for the nuclear mass excess, shown for the test (upper), extpl (middle), and new (lower) datasets. The vertical dotted line at zero marks perfect agreement between predicted and experimental values.

from the extpl and new sets, turning green in the upper panel. This is particularly important for the extrapolated and newly measured set of nuclei, as their improved agreement with experiment demonstrates that the GBR correction generalizes well beyond the training region. It is further seen

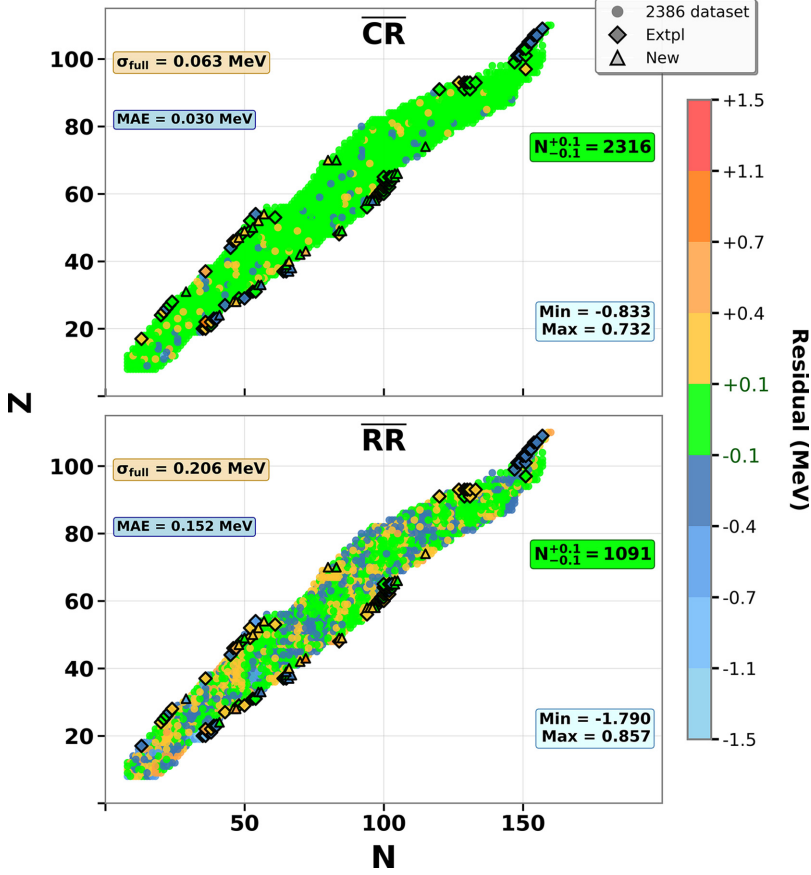


FIG. 4. Values of the weighted-average nuclear mass excess in the N - Z plane. The lower and upper panels show the results obtained using raw (RR) and CR, respectively. The notation $N_{-0.1}^{+0.1}$ denotes the number of nuclei whose predicted nuclear mass excess has an absolute error of less than 100 keV (green symbols). The values of σ_{full} and MAE are also indicated in each panel. The residual range for RR (-1.790 to $+0.857$ MeV) is markedly asymmetric, whereas after GBR correction the CR range (-0.833 to $+0.732$ MeV) becomes nearly symmetric about zero, reflecting the suppression of overbinding bias.

that the spread of $\delta\bar{M}^{\text{RR}}$ is markedly asymmetric, with the minimum and maximum values indicating a tendency toward overbinding in the raw residuals. After GBR correction, this asymmetry is reduced and the range of $\delta\bar{M}^{\text{CR}}$ becomes nearly symmetric about zero, reflecting that the GBR correction not only reduces the magnitude of residuals but also suppresses the directional bias present in the raw predictions. The values of σ_{full} and mean absolute error (MAE) for the CR are reduced by more than a factor of 3 and 6, respectively, relative to the RR, and the maximum spread of the residuals also narrows considerably.

The advantage of the weighted-averaging approach in improving nuclear mass-excess predictions is evident from Table II and Figs. 2 and 4. These effects are further illustrated in Fig. 5, which compares the weighted average of corrected residuals for a representative set of 15 nuclei, selected over a wide mass range, with those from individual nuclear mass models obtained using the GBR. By definition, unbiased residuals exhibit both positive and negative deviations and, under unbiased model assumptions, follow a distribution centered at zero with finite variance [47]. This statistical symmetry implies that averaging residuals leads to partial cancellation of deviations, thereby reducing the net error. Consistent with this, the corrected residuals from individual models are randomly

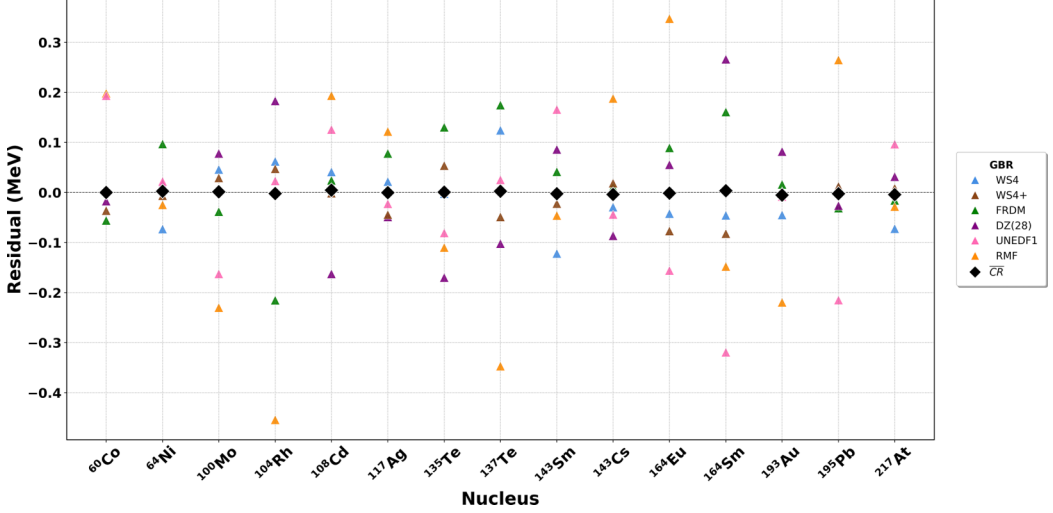


FIG. 5. Comparison of the weighted-average nuclear mass excess values, obtained using corrected residuals, with those predicted individually by each nuclear mass model employing GBR method for a selected set of nuclei. The magnitude of the weighted-average CR is considerably smaller than those for the individual models, primarily due to partial cancellation of model-specific deviations.

scattered close to zero, often with opposite signs. It may also be noted that out of six different values of corrected residuals, for each of the 15 nuclei considered, we have roughly equal number of points on either side of zero line leading to partial cancellation of the residuals while performing the weighted average. This partial cancellation brings the weighted-average residuals, shown as black diamonds, consistently closer to zero than those of any single model. Models that overestimate the mass excess ($\delta\bar{M}^{\text{CR}} < 0$) for certain isotopes are balanced by others that underestimate them ($\delta\bar{M}^{\text{CR}} > 0$), reducing overall uncertainty. Because weights are assigned according to each model's predictive accuracy on the test set, better-performing models exert greater influence on the final estimate, enhancing the stability of ensemble predictions. This mutual error balancing not only lowers RMSE values (Table II) but also yields more robust and reliable predictions across the nuclear chart, including extrapolative and data-sparse regions where individual models typically underperform, as seen in Table II and Fig. 5.

A. Robustness of ELMA

To test the robustness of the ELMA framework, we evaluated its performance under different train-test data splits, as summarized in Table III. The training RMSE remains nearly unchanged across all splits, indicating stable learning behavior. For the test, extpl, and full datasets, the RMSE decreases marginally as the training fraction increases, reflecting the expected benefit of larger training statistics. The new dataset shows negligible variation across all splits, demonstrating that the extrapolative performance of ELMA is robust and largely insensitive to the choice of partition. This trend is consistent with previous machine-learning-based nuclear mass studies [28]. Based on this analysis, the 85/15 train-test split is adopted for all calculations, as it offers the best balance between training statistics and validation coverage.

The choice of train-test partition introduces an element of randomness that may influence the reported RMSE values. To quantify this uncertainty, the entire training and evaluation procedure has been repeated over 5000 independent random train-test splits, and the resulting RMSE distribution for the CR weighted average have examined via probability distribution analysis, as shown in Fig. 6. For the test dataset, the CR distribution peaks near 0.131 MeV with a spread of ~ 0.0102

TABLE III. Values of RMSE (σ , in MeV) obtained using ELMA for different train-test data splits, reported for the training set (train), test set (test), extrapolated nuclei (extpl), the complete AME2020 dataset (2020), nuclei reported for the first time after AME2020 (new), and the full dataset. The percentages indicate the proportion of data allocated to training and testing, respectively.

Dataset split (%)	σ (MeV)					
	Train	Test	Extpl	2020	New	Full
70/30	0.0140	0.1310	0.1880	0.0784	0.2088	0.0814
75/25	0.0151	0.1307	0.1889	0.0731	0.2031	0.0761
80/20	0.0166	0.1271	0.1804	0.0656	0.2125	0.0693
85/15	0.0179	0.1245	0.1816	0.0590	0.2141	0.0633

MeV, confirming that the GBR correction remains stable across different partitions. For the extpl dataset, the CR distribution is sharply peaked near 0.183 MeV with a spread of ~ 0.009 MeV, indicating highly reproducible predictions regardless of the chosen split. For the new dataset, the CR distribution peaks near 0.205 MeV with a spread of ~ 0.014 MeV, reflecting the greater variability inherent in this smaller and more challenging set of nuclei. Across all three datasets, the CR distributions are consistently narrow and well-peaked, confirming that the GBR correction yields stable and reproducible predictions across different data partitions. Taken together with the split-comparison results of Table III, these observations establish that the predictions of ELMA are statistically robust, independent of how the training and test data are partitioned.

B. Application of ELMA

The refinement of nuclear mass predictions through the use of ELMA may provide a foundation for broader applications in nuclear physics and astrophysics. Having minimized the deviations and enhanced predictability across the nuclear chart, the resulting mass table becomes a more reliable resource for theoretical and experimental studies alike. In nuclear structure and decay studies, accurate nuclear masses serve as a cornerstone for calculating nucleon separation energies, decay energies, and reaction thresholds. Even small improvements in mass prediction accuracy can lead to substantial gains in the reliability of theoretical models and experimental planning.

One of the most direct and critical applications of an improved mass table is the determination of the Q_α values, which represent the difference in mass excess or binding energies of the parent and daughter nuclei. These values are required for determining the properties of alpha decay, assessing the boundaries of nuclear stability, and interpreting decay chains of newly synthesized heavy and superheavy elements. For superheavy nuclei in particular, where experimental mass measurements are scarce or absent, the precision of the predicted Q_α values is often considered a deciding factor in confirming the discovery of new isotopes.

The Q_α values in this analysis are calculated directly from the AME2020 experimental mass-excess database using experimentally known parent-daughter pairs. The resulting dataset contains 2321 Q_α pairs spanning nuclei with proton numbers in the range $10 \leq Z \leq 110$. Among these, 1912 parent-daughter pairs belong to the training set, 340 belong to the test set, and the remaining 69 correspond to the independent extrapolation datasets.

In Fig. 7, we plot residuals of the Q_α values calculated with weighted-averaged values of RR and CR for nuclear mass excess. The refinement in nuclear mass prediction is evident, since the residual distribution of Q_α is narrower as can be visualized from figure by enhancement in the number of green symbols (upper panel). The RMSE reduces to 0.081 MeV (upper panel), which is a factor of 2 smaller than those for the RR (lower panel). In addition, the number of nuclei for which $|\delta Q_\alpha| < 0.1$ MeV increases from 1086 to 2089, reflecting a substantial enhancement in predictive accuracy of the Q_α values across a wide range of the periodic chart. This enhancement in predictive

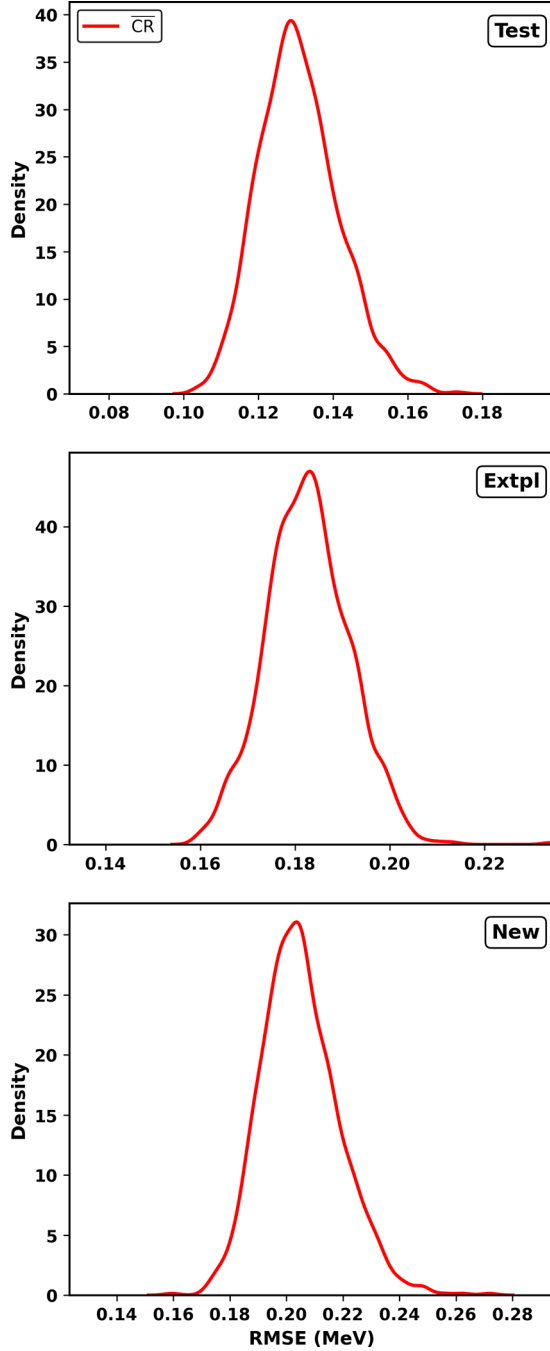


FIG. 6. The weighted-average values of RMSE for corrected residuals ($\overline{CR}$, red solid) obtained from 5000 independent random train-test splits, shown for the test (upper), extpl (middle), and new (lower) datasets.

accuracy is particularly critical, as an uncertainty of just ± 0.1 MeV in Q_α demonstrates a profound sensitivity of decay rates, where an increase of that magnitude may result in a predicted half-life changed by $\sim 40\%$ of its original value.

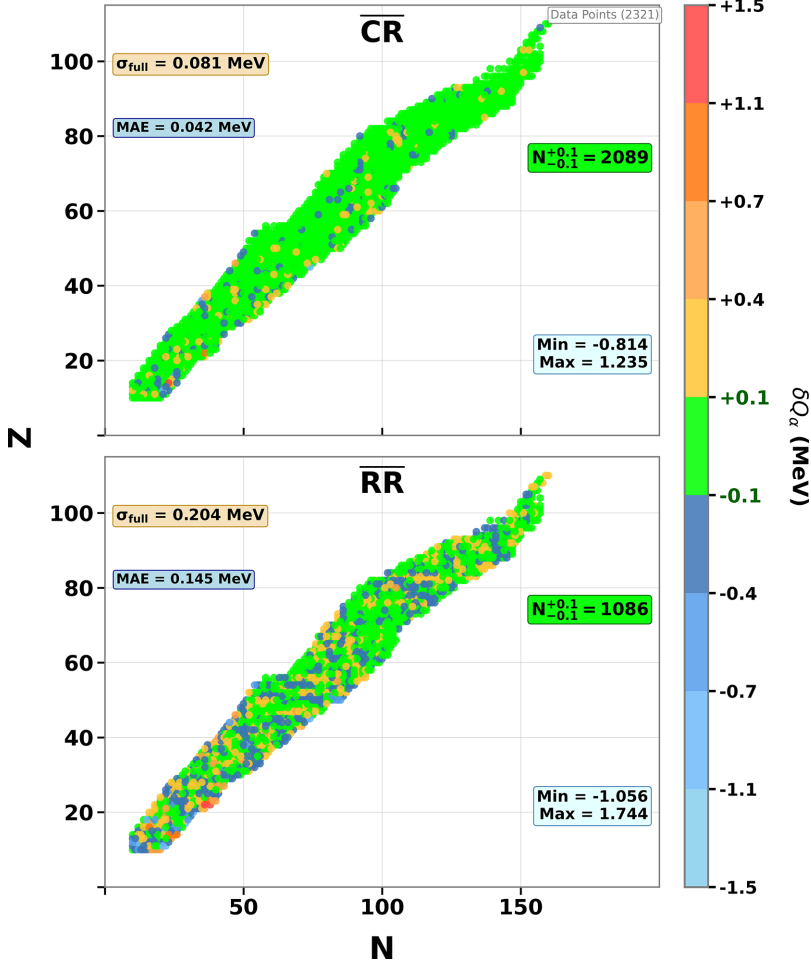


FIG. 7. Residuals for α -decay Q values obtained using the weighted-average values of raw residuals (lower panel) and corrected residuals (upper panel) of the nuclear mass excess. The values of σ and $N_{-0.1}^{+0.1}$ are also given; $N_{-0.1}^{+0.1}$ indicates the number of cases with $|\delta Q_\alpha| < 100$ keV (green symbols).

C. Nuclear mass table for astrophysical applications

The values of mass excess and binding energies obtained using nuclear mass model ELMA are calculated for 6303 nuclei. This mass table is made available as Supplemental Material [48]. By providing more accurate mass predictions, this table can be used to integrate into a variety of research workflows, including r -process nucleosynthesis, β -decay-energy calculations, and the evaluation of reaction Q values in astrophysical environments. Furthermore, its enhanced precision in Q_α estimation makes it as a valuable resource for guiding experimental results in the synthesis and identification of isotopes at the limits of the nuclear chart. We have also deployed ELMA on a web interface, accessible via the link provided in Ref. [49]. In the future, we plan to continuously update ELMA by incorporating new experimental measurements as well as theoretical advancements, thereby enhancing its predictive accuracy.

V. CONCLUSIONS

We employed ensemble learning with model averaging (ELMA) to improve the predictive accuracy of global nuclear mass models. The framework combines GBR-based residual corrections from six theoretically distinct mass models representing macroscopic-microscopic, nonrelativistic self-consistent, and relativistic mean-field approaches. The corrected residuals are subsequently integrated through a weighted average to construct a unified nuclear mass model.

The principal novelty of ELMA lies in combining machine-learned residual corrections from multiple theoretical models within a single ensemble framework, thereby exploiting complementary model-dependent systematics rather than relying on a single corrected mass model. The resulting conflated mass model achieves an RMSE of approximately 0.0571 MeV for the AME2020 dataset. The model’s reliability in extrapolative regions is confirmed by its high precision on an independent test set of extrapolated and newly measured nuclei. The stability of the framework is further confirmed through 5000 independent random train-test splits, which produce consistently narrow RMSE distributions, indicating that the results are robust against variations in the data partition.

The improved mass predictions propagate directly to derived observables, leading to substantially improved predictions of Q_α values in comparison with experimental data. Using ELMA, we have generated a database of nuclear mass excesses and binding energies for approximately 6300 nuclei. The database is publicly available through the web interface. Reference [49] providing a dynamic resource for applications including r -process nucleosynthesis, decay-energy calculations, and reaction Q -value evaluations. Future work will focus on extending the framework to additional nuclear observables.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Jatin Fulwani for his valuable assistance in developing the ELMA web interface. N.C. and G.S. express gratitude for the support received from the Department of Science & Technology (DST), Government of Rajasthan, Jaipur, India under F24(1)DST/R&D/2024-EAC-00378-6549873/819. T.G. and N.P. acknowledge support from the Croatian Science Foundation under the project Relativistic Nuclear Many-Body Theory in the Multimessenger Observation Era (HRZZ-IP-2022-10-7773). B.K.A. acknowledges with gratitude the support received under the Raja Ramanna Chair scheme of the Department of Atomic Energy (DAE), Government of India. N.P. acknowledges support from the project “Implementation of cutting-edge research and its application as part of the Scientific Center of Excellence for Quantum and Complex Systems, and Representations of Lie Algebras,” Grant No. PK.1.1.10.0004, co-financed by the European Union through the European Regional Development Fund—Competitiveness and Cohesion Programme 2021–2027.

DATA AVAILABILITY

The data that support the findings of this article are openly available at [49].

-
- [1] D. Martin, A. Arcones, W. Nazarewicz, and E. Olsen, Impact of nuclear mass uncertainties on the r process, *Phys. Rev. Lett.* **116**, 121101 (2016).
 - [2] M. Amould, S. Goriely, and K. Takahashi, The r -process of stellar nucleosynthesis: Astrophysics and nuclear physics achievements and mysteries, *Phys. Rep.* **450**, 97 (2007).
 - [3] D. Lunney and I. Collaboration, Extending and refining the nuclear mass surface with ISOLTRAP, *J. Phys. G: Nucl. Part. Phys.* **44**, 064008 (2017).
 - [4] M. Smith, M. Brodeur, T. Brunner, S. Ettenauer, A. Lapierre, R. Ringle, V. L. Ryjkov, F. Ames, P. Bricault, G. W. F. Drake, P. Delheij, D. Lunney, F. Sarazin, and J. Dilling, First Penning-trap mass measurement of the exotic halo nucleus ^{11}Li , *Phys. Rev. Lett.* **101**, 202501 (2008).

- [5] M. Hausmann *et al.*, Isochronous mass measurements of hot exotic nuclei, [Hyperfine Interact.](#) **132**, 289 (2001).
- [6] T. Yamaguchi, Y. Yamaguchi, and A. Ozawa, The challenge of precision mass measurements of short-lived exotic nuclei: Development of a new storage ring mass spectrometry, [Int. J. Mass Spectrom.](#) **349–350**, 240 (2013).
- [7] A. Czihaly, S. Beck, J. Bergmann, C. L. Brown, T. Brunner, T. Dickel, J. Dilling, E. Dunling, J. Flowerdew, D. Fusco, *et al.*, Exploring the nuclear chart via precision mass spectrometry with the TITAN MR-TOF MS, [Atoms](#) **13**, 6 (2025).
- [8] P. F. Giesel, M. Benhatchi, A. Arcones, D. Atanasov, K. Blaum, R. B. Cakirli, J. Kartheim, D. Lange, Y. A. Litvinov, D. Lunney, V. Manea, M. Mougeot, S. Naimi, L. Nies, C. Schweiger, L. Schweikhard, and F. Wienholtz, *Mass measurements of neutron-rich Ag and In isotopes for r-process nucleosynthesis studies*, Technical Report No. CERN-INTC-2024-061, INTC-P-716 (CERN, Geneva, 2024).
- [9] A. Boehnlein, M. Diefenthaler, N. Sato, M. Schram, V. Ziegler, C. Fanelli, M. Hjorth-Jensen, T. Horn, M. P. Kuchera, D. Lee, W. Nazarewicz, P. Ostroumov, K. Orginos, A. Poon, X.-N. Wang, A. Scheinker, M. S. Smith, and L.-G. Pang, *Colloquium: Machine learning in nuclear physics*, [Rev. Mod. Phys.](#) **94**, 031003 (2022).
- [10] W. J. Huang, M. Wang, F. G. Kondev, G. Audi, and S. Naimi, The AME 2020 atomic mass evaluation (I). Evaluation of input data, and adjustment procedures, [Chin. Phys. C](#) **45**, 030002 (2021).
- [11] G. Audi, The history of nuclidic masses and of their evaluation, [Int. J. Mass Spectrom.](#) **251**, 85 (2006).
- [12] A. H. Wapstra, G. Audi, and C. Thibault, The Ame2003 atomic mass evaluation, [Nucl. Phys. A](#) **729**, 129 (2003).
- [13] H. A. Bethe and R. F. Bacher, Nuclear Physics A. Stationary states of nuclei, [Rev. Mod. Phys.](#) **8**, 82 (1936).
- [14] A. Bhagwat, Simple nuclear mass formula, [Phys. Rev. C](#) **90**, 064306 (2014).
- [15] 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).
- [16] J. Duflo and A. P. Zuker, Microscopic mass formulas, [Phys. Rev. C](#) **52**, R23 (1995).
- [17] N. Wang, M. Liu, X. Wu, and J. Meng, Surface diffuseness correction in global mass formula, [Phys. Lett. B](#) **734**, 215 (2014).
- [18] N. N. Ma, H. F. Zhang, X. J. Bao, and H. F. Zhang, Basic characteristics of nuclear landscape by improved Weizsäcker-Skyrme-type nuclear mass model, [Chin. Phys. C](#) **43**, 044105 (2019).
- [19] Y. Aboussir, J. M. Pearson, A. K. Dutta, and F. Tondeur, Nuclear mass formula via an approximation to the Hartree–Fock method, [At. Data Nucl. Data Tables](#) **61**, 127 (1995).
- [20] S. Goriely, N. Chamel, and J. M. Pearson, Further explorations of Skyrme–Hartree–Fock–Bogoliubov mass formulas. XIII. The 2012 atomic mass evaluation and the symmetry coefficient, [Phys. Rev. C](#) **88**, 024308 (2013).
- [21] D. Vretenar, A. V. Afanasjev, G. A. Lalazissis, and P. Ring, Relativistic Hartree–Bogoliubov theory: Static and dynamic aspects of exotic nuclear structure, [Phys. Rep.](#) **409**, 101 (2005).
- [22] S.-G. Zhou, Multidimensionally constrained covariant density functional theories—Nuclear shapes and potential energy surfaces, [Phys. Scr.](#) **91**, 063008 (2016).
- [23] P. Ring, Relativistic mean field theory in finite nuclei, [Prog. Part. Nucl. Phys.](#) **37**, 193 (1996).
- [24] J. S. Zheng, N. Y. Wang, Z. Y. Wang, Z. M. Niu, Y. F. Niu, and B. Sun, Mass predictions of the relativistic mean-field model with the radial basis function approach, [Phys. Rev. C](#) **90**, 014303 (2014).
- [25] R. Utama and J. Piekarewicz, Refining mass formulas for astrophysical applications: A Bayesian neural network approach, [Phys. Rev. C](#) **96**, 044308 (2017).
- [26] 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).
- [27] R. Utama, J. Piekarewicz, and H. Prosper, Nuclear mass predictions for the crustal composition of neutron stars: A Bayesian neural network approach, [Phys. Rev. C](#) **93**, 014311 (2016).
- [28] E. Yüksel, D. Soydaner, and H. Bahtiyar, Nuclear mass predictions using machine learning models, [Phys. Rev. C](#) **109**, 064322 (2024).

- [29] Z. Niu and H. Liang, Nuclear mass predictions with machine learning reaching the accuracy required by r -process studies, *Phys. Rev. C* **106**, L021303 (2022).
- [30] Y. Huang, J. Chen, J. Jia, L. M. Liu, Y. G. Ma, and C. Zhang, Validation and extrapolation of atomic masses with a physics-informed fully connected neural network, *Phys. Rev. C* **111**, 034329 (2025).
- [31] G. P. Liu, H. L. Wang, Z. Z. Zhang, and M. L. Liu, Model-repair capabilities of tree-based machine-learning algorithms applied to theoretical nuclear mass models, *Phys. Rev. C* **111**, 024306 (2025).
- [32] W. Ye and N. Wan, Understanding on prediction differences among theoretical mass models with machine learning techniques, *Phys. Rev. C* **111**, 044317 (2025).
- [33] Y. Saito, I. Dillmann, R. Kruecken, M. R. Mumpower, and R. Surman, Uncertainty quantification of mass models using ensemble Bayesian model averaging, *Phys. Rev. C* **109**, 054301 (2024).
- [34] V. Kejzlar, L. Neufcourt, and W. Nazarewicz, Local Bayesian Dirichlet mixing of imperfect models, *Sci. Rep.* **13**, 19600 (2023).
- [35] E. Alhassan, D. Rochman, G. Schnabel, and A. J. Koning, Bayesian model averaging (BMA) for nuclear data evaluation, *Nucl. Sci. Tech.* **35**, 205 (2024).
- [36] S. P. Adam, S.-A. N. Alexandropoulos, P. M. Pardalos, and M. N. Vrahatis, No free lunch theorem: A review, in *Approximation and Optimization: Algorithms, Complexity and Applications*, edited by I. C. Demetriou and P. M. Pardalos (Springer, Cham, 2019), pp. 57–82.
- [37] L.-G. Pang, Studying high-energy nuclear physics with machine learning, *Int. J. Mod. Phys. E* **33**, 2430009 (2024).
- [38] I. Bentley, J. Tedder, M. Gebran, and A. Paul, High precision binding energies from physical-feature-instructed machine learning, *Phys. Rev. C* **111**, 034305 (2025).
- [39] J. H. Friedman, Greedy function approximation: A gradient boosting machine, *Ann. Stat.* **29**, 1189 (2001).
- [40] Z. P. Gao, Y. J. Wang, H. L. Lü, Q. F. Li, C. W. Shen, and L. Liu, Machine learning the nuclear mass, *Nucl. Sci. Tech.* **32**, 109 (2021).
- [41] L. W. Rizkallah, Enhancing the performance of gradient boosting trees on regression problems, *J. Big Data* **12**, 35 (2025).
- [42] J. H. Friedman, Stochastic gradient boosting, *Comput. Stat. Data Anal.* **38**, 367 (2002).
- [43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.* **12**, 2825 (2011).
- [44] M. Kortelainen, J. McDonnell, W. Nazarewicz, P.-G. Reinhard, J. Sarich, N. Schunck, M. Stoitsov, and S. Wild, Nuclear energy density optimization: Large deformations, *Phys. Rev. C* **85**, 024304 (2012).
- [45] L. S. Geng, H. Toki, S. Sugimoto, and J. Meng, Relativistic mean field theory for deformed nuclei with pairing correlations, *Prog. Theor. Phys.* **110**, 921 (2003).
- [46] L. S. Geng, *Ground state properties of finite nuclei in the relativistic mean field model*, Ph.D. thesis, Research Center for Nuclear Physics (RCNP), Osaka University, 2005.
- [47] D. C. Montgomery, E. A. Peck, and G. G. Vining, Model adequacy checking, in *Introduction to Linear Regression Analysis*, 5th ed. (John Wiley & Sons, New Jersey, 2021), Chap. 4, pp. 134–169.
- [48] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/8n1w-f3b7> for the complete ELMA nuclear mass table.
- [49] <https://ddnp.in>.