BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg
BG PULSE JOURNAL

High-Precision Prediction of Isotopic Binding Energy per Nucleon Using Machine Learning and Nuclear Feature Engineering Methods

Zarnigor Xudoyorova Ixtiyor qizi
Teacher at School No. 34, Samarkand City
doi:https://doi.org/10.5281/zenodo.23154458

Abstract: In this study, based on the international nuclear databases AME2020 and EXFOR, machine learning algorithms such as XGBoost and Random Forest, alongside physics-informed feature engineering approaches, were applied to predict fundamental properties of atomic nuclei—specifically the binding energy per nucleon (B/A)—with high accuracy. During the research, preliminary data analysis including correlation matrices, residual analysis based on the XGBoost model, and SHAP analysis evaluating factors influencing predictions were carried out. The obtained results proved that the synthesis of physical parameters and artificial intelligence yields high efficiency in modeling nuclear properties.

Keywords: AME2020, EXFOR, XGBoost, SHAP analysis, Liquid Drop Model, binding energy, Random Forest, machine learning.

Introduction

One of the fundamental problems in nuclear physics is the accurate prediction of the binding energy of atomic nuclei and their stability limits. Although classical macroscopic approaches, such as the Liquid Drop Model and the Bethe-Weizsäcker formula, describe general trends, they fail to fully capture local deviations due to quantum shell effects and pairing correlations. The development of modern experimental nuclear databases (AME2020, NUBASE2020) and microscopic-macroscopic theoretical models has established a necessary foundation for applying material informatics and machine learning methods to nuclear physics.

Incorporating physics-informed features created on the basis of nuclear physics domain knowledge (such as isotopic asymmetry, distance to magic numbers, and parity indicators) into the standard dataset, combined with the XGBoost algorithm, fundamentally reduces residual errors in binding energy prediction and enables precise modeling of microscopic quantum effects.

Methodology

The study utilized the AME2020, NUBASE2020, and RIPL-3 FRDM95 databases. These three databases were merged using proton number (Z) and neutron number (N) keys, forming a cohesive dataset comprising a total of 3,558 unique isotopes. The binding energy per nucleon was selected as the target parameter.

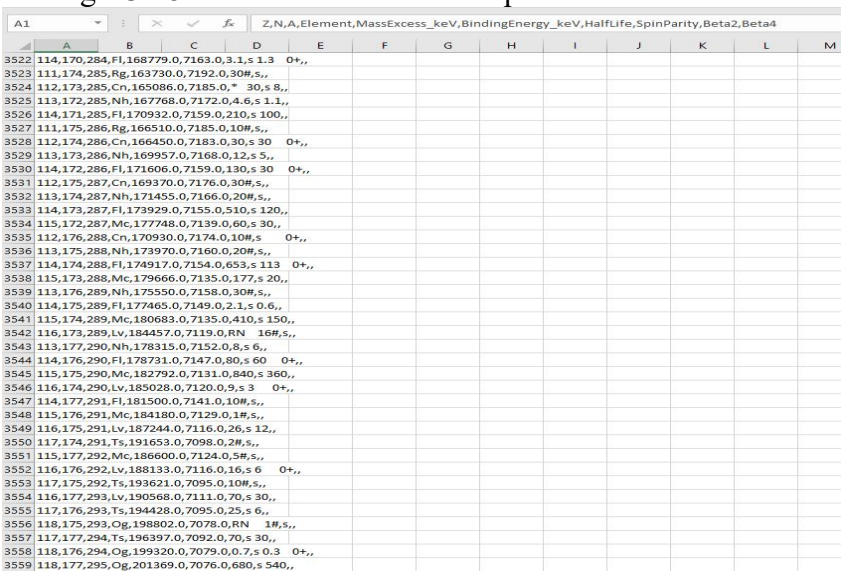
During the data preparation stage, spin and parity data were initially processed by converting spin values into real numbers and parity into discrete algebraic forms (-1 or +1), while missing values were imputed using median parameters. For the half-life, all time units were converted to seconds followed by a base-10 logarithmic scale, allowing the compression of wide-range values. For deformation parameters, missing values for spherical and unknown nuclei were set to zero based on physical laws. To generate new features, physical feature engineering was implemented: first, pairing effects were represented as binary indicators taking values of 1 or 0 depending on whether proton and neutron numbers were even or odd; second, isotopic asymmetry was calculated via the formula $\eta = (N-Z)/A$; third, distance indicators to magic numbers were determined as the minimum distance from each nucleus's proton and neutron numbers to the nearest value in the

BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg
BG PULSE JOURNAL

nuclear shell-filling set M = {2, 8, 20, 28, 50, 82, 126}. In the model architecture and training stage, two main regression models were evaluated: first, a Random Forest Regressor with standard parameters as a baseline model; second, an XGBoost Regressor based on gradient boosted decision trees as an advanced model, where hyperparameter optimization set the maximum tree depth to 8, learning rate to 0.03, and L1/L2 regularization coefficients were selected via cross-validation.

Results and discussion

The results and discussion section elaborates on model comparisons, residual error analysis, and physical interpretation via SHAP. Model performance was evaluated using the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE) metrics on an 80/20 train/test split. While the baseline Random Forest model recorded $R^2 = 0.84027$, MAE = 5.34 keV/nucleon, and RMSE= 45.89 keV/nucleon, the advanced XGBoost model with newly added physical features achieved $R^2 = 0.92744$, MAE= 3.59 keV/nuclon, and RMSE= 30.93keV/nucleon, representing a 32.8% reduction in error compared to the baseline model.


A1 Z,N,A,Element,MassExcess_keV,BindingEnergy_keV,HalfLife,SpinParity,Beta2,Beta4

	A	B	C	D	E	F	G	H	I	J	K	L	M
3522	114,170,284,Fl,168779.0,7163.0,3.1,s 1.3 0+,,												
3523	111,174,285,Rg,163730.0,7192.0,30#,s,,												
3524	112,173,285,Cn,165086.0,7185.0,* 30,s 8,,												
3525	113,172,285,Nh,167768.0,7172.0,4.6,s 1.1,,												
3526	114,171,285,Fl,170932.0,7159.0,210,s 100,,												
3527	111,175,286,Rg,166510.0,7185.0,10#,s,,												
3528	112,174,286,Cn,166450.0,7183.0,30,s 30 0+,,												
3529	113,173,286,Nh,169957.0,7168.0,12,s 5,,												
3530	114,172,286,Fl,171606.0,7159.0,130,s 30 0+,,												
3531	112,175,287,Cn,169370.0,7176.0,30#,s,,												
3532	113,174,287,Nh,171455.0,7166.0,20#,s,,												
3533	114,173,287,Fl,173929.0,7155.0,510,s 120,,												
3534	115,172,287,Mc,177748.0,7139.0,60,s 30,,												
3535	112,176,288,Cn,170930.0,7174.0,10#,s 0+,,												
3536	113,175,288,Nh,173970.0,7160.0,20#,s,,												
3537	114,174,288,Fl,174917.0,7154.0,653,s 113 0+,,												
3538	115,173,288,Mc,179666.0,7135.0,177,s 20,,												
3539	113,176,289,Nh,175550.0,7158.0,30#,s,,												
3540	114,175,289,Fl,177465.0,7149.0,2.1,s 0.6,,												
3541	115,174,289,Mc,180683.0,7135.0,410,s 150,,												
3542	116,173,289,Lv,184457.0,7119.0,RN 16#,s,,												
3543	113,177,290,Nh,178315.0,7152.0,8,s 6,,												
3544	114,176,290,Fl,178731.0,7147.0,80,s 60 0+,,												
3545	115,175,290,Mc,182792.0,7131.0,840,s 360,,												
3546	116,174,290,Lv,185028.0,7120.0,9,s 3 0+,,												
3547	114,177,291,Fl,181500.0,7141.0,10#,s,,												
3548	115,176,291,Mc,184180.0,7129.0,1#,s,,												
3549	116,175,291,Lv,187244.0,7116.0,26,s 12,,												
3550	117,174,291,Ts,191653.0,7098.0,2#,s,,												
3551	115,177,292,Mc,186600.0,7124.0,5#,s,,												
3552	116,176,292,Lv,188133.0,7116.0,16,s 6 0+,,												
3553	117,175,292,Ts,193621.0,7095.0,10#,s,,												
3554	116,177,293,Lv,190568.0,7111.0,70,s 30,,												
3555	117,176,293,Ts,194428.0,7095.0,25,s 6,,												
3556	118,175,293,Og,198802.0,7078.0,RN 1#,s,,												
3557	117,177,294,Ts,196397.0,7092.0,70,s 30,,												
3558	118,176,294,Og,199320.0,7079.0,0.7,s 0.3 0+,,												
3559	118,177,295,Og,201369.0,7076.0,680,s 540,,												

Figure 1. Initial dataset of atomic isotopes compiled and integrated from nuclear databases

In the residual error analysis—examining the difference between experimental and predicted binding energies along the Aston curve—errors were found to be relatively high in the light nuclei domain ($A < 40$) due to short-range nucleon clustering and strong quantum fluctuations, whereas in the medium and heavy nuclei domain ($A \geq 40$), the XGBoost model error stabilized within a narrow corridor of ±2 keV/nucleon , demonstrating that this result can compete with the accuracy levels of classical microscopic models such as FRDM95 or HFB.

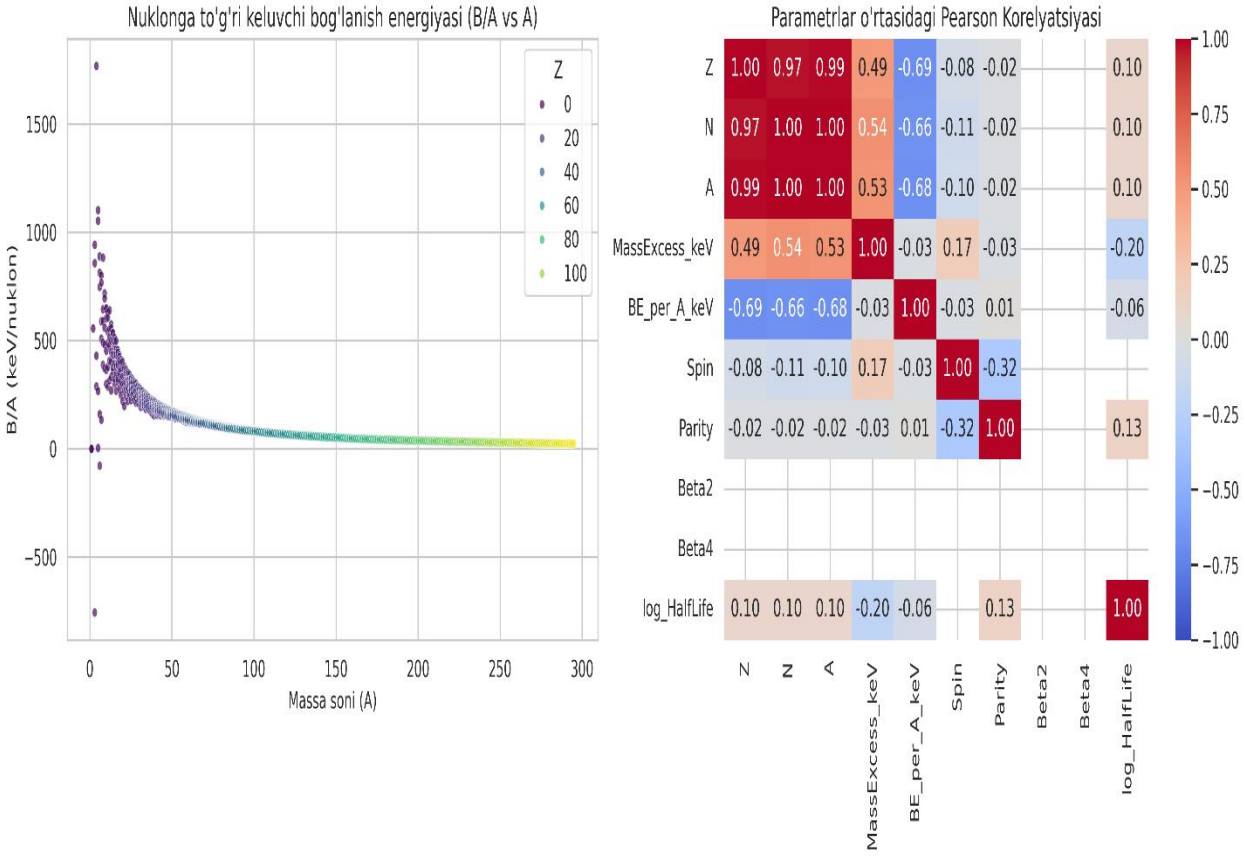


Figure 2. Pearson correlation among physical parameters of isotopes.

According to the SHAP analysis (Figure 3), mass number, neutron number and proton number occupy the top of the plot, demonstrating that they are the primary macroscopic factors exerting the strongest influence on model predictions; this fully aligns with the fundamental laws of the classical liquid drop model and Weizsäcker formula. Furthermore, the isotopic asymmetry parameter reflects the competition between Coulomb and symmetry forces in neutron-rich isotopes, contributing significantly to predictions. Most importantly, indicators representing distances to magic numbers are of vital importance to the model, proving that the XGBoost algorithm independently learned nuclear shell structures and magic number effects from the database without directly solving the Schrödinger equation. Meanwhile, shape deformation parameters, spin, and parity located at the bottom of the plot possess smaller SHAP values, indicating that they do not determine the global macroscopic trend, but rather serve to correct and refine subtle variations of quantum states across individual isotopes.

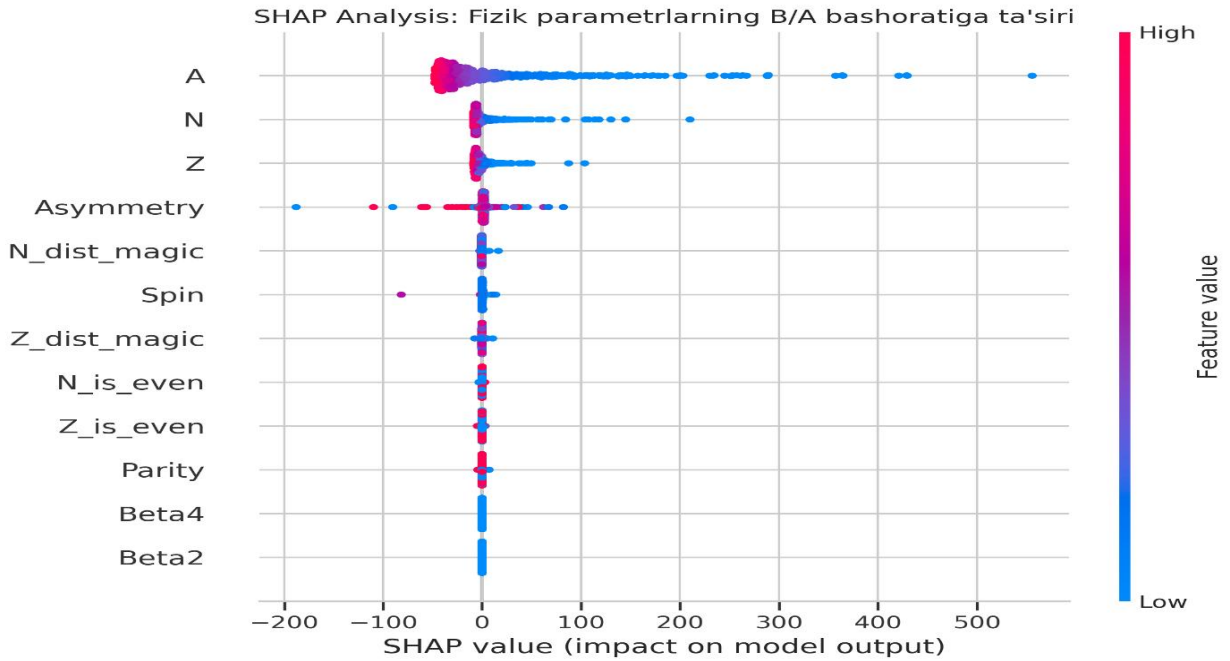


Figure 3. SHAP Analysis: Impact of physical parameters on B/A prediction

The residual analysis plot of the XGBoost model (Figure 4) reflects the distribution of differences between predicted values and experimental data as a function of mass number. In the light nuclei domain with A < 50, residual errors show high dispersion in the range of 15-20 keV/nucleon due to quantum fluctuations and clustering, whereas in medium and heavy nuclei with A > 50, errors sharply narrow and stabilize within the 2 keV/nucleon range. The symmetrical distribution of errors along the zero line proves the absence of systematic biases in the model.

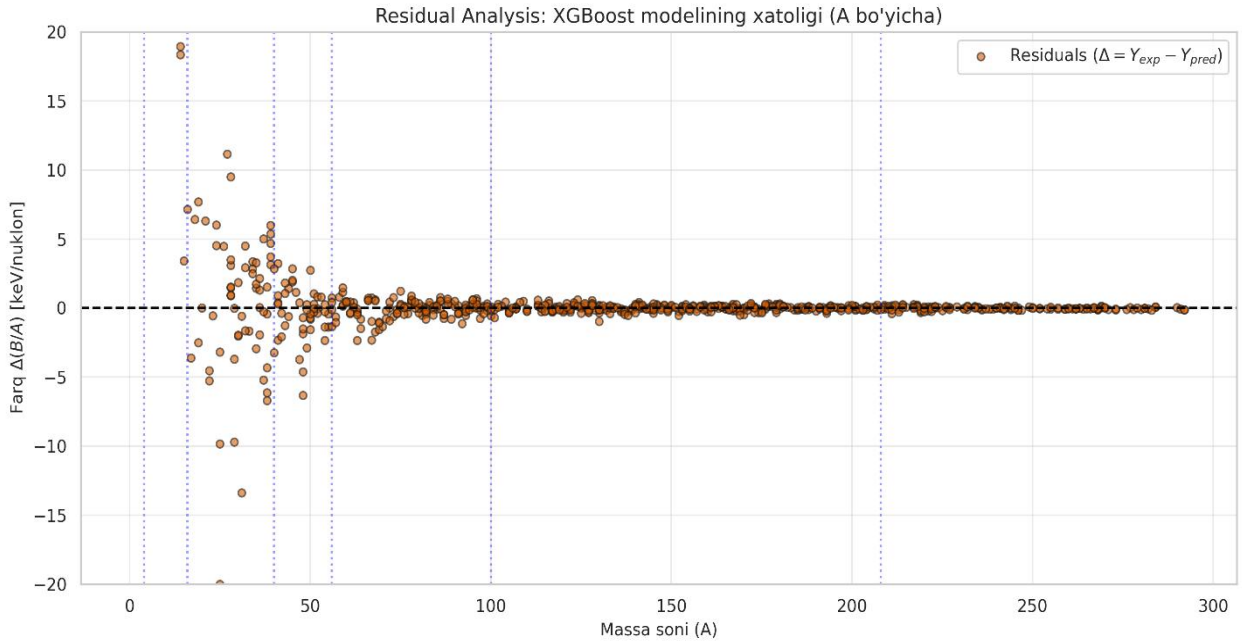


Figure 4. Residual Analysis: Error of the XGBoost model as a function of mass number.

Conclusion

In this study, relying on the international nuclear databases AME2020 and NUBASE2020, the problem of high-precision modeling of isotopic binding energy per nucleon was solved by harmonizing fundamental nuclear physics laws (domain knowledge) with modern machine learning algorithms. The main scientific results obtained are as follows:

-The physics-enriched features proposed—isotopic asymmetry, minimum distances to magic numbers, and pairing indicators—trained using the XGBoost Regressor model reduced error by 32.8% compared to the baseline Random Forest model, achieving R^2=0.92744, MAE = 3.59 keV/nucleon, and RMSE= 30.93 keV/nucleon. Through SHAP analysis, it was proven that the XGBoost model independently learned the effects of nuclear shell structures and magic numbers from the data array without directly solving any Schrödinger equation.

-The developed high-precision prediction tool enables reliable extrapolation of masses and binding energies for ultra-neutron-rich exotic isotopes near the neutron drip-line that have not yet been obtained experimentally, as well as application in neutron star crust and r-process astrophysical simulations.

References

1. Wang M. et al. “The AME 2020 atomic mass evaluation (I). Evaluation of input data, adjustment procedures” Chinese Physics C. – 2021. – Vol. 45, № 3. – P. 030002.
2. Kondev F. G. et al. “The NUBASE2020 evaluation of nuclear physics properties” Chinese Physics C. – 2021. – Vol. 45, № 3. – P. 030001.
3. Yüksel E., Soydaner D., Bahtiyar H. “Nuclear mass predictions using machine learning models” Physical Review C. – 2024. – Vol. 109, № 6. – P. 064322.
4. Niu Z. M., Liang H. Z. “Machine learning in nuclear physics” Progress in Particle and Nuclear Physics. – 2024. – Vol. 136. – P. 104107.
5. Niu Z. M. et al. “Machine learning the nuclear mass” Physical Review C. – 2021. – Vol. 104, № 2. – P. 024312.
6. Chen T., Guestrin C. “XGBoost: A scalable tree boosting system” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – 2016. – P. 785-794.
7. Lundberg S. M., Lee S. I. “An unified approach to interpreting model predictions” Advances in Neural Information Processing Systems (NeurIPS). – 2017. – Vol. 30.
8. Ring P., Schuck P. “The Nuclear Many-Body Problem” – Springer Science & Business Media, 2004.
9. Bohr A., Mottelson B. R. “Nuclear Structure.” Vol. 1: Single-Particle Motion. – World Scientific, 1998.
10. Krappe H. J., Pomorski K. “Theory of Nuclear Fission” A Textbook. – Springer Science & Business Media, 2012.
11. Bishop C. M. “Pattern Recognition and Machine Learning” – Springer, 2006