Predicting Nuclear Binding Energy Using Generalized Additive Model

Kristiyan Laoli[1⋆]

1 Department of Physics, Universitas Gadjah Mada, Yogyakarta, Indonesia

⋆ kristiyanlaoli@mail.ugm.ac.id

Abstract

This study presents the implementation of the Generalized Additive Model (GAM) for predicting nuclear binding energies. GAM, a non-parametric regression model, effectively captures complex, non-linear relationships between input variables and the output, providing an interpretable framework for understanding the contribution of each nuclear property to the binding energy. The model's performance is evaluated using data from the Atomic Mass Evaluation (AME) 2020, yielding promising results with a Root Mean Square Error (RMSE) of 0.3 MeV. This study demonstrates the potential of machine learning methods, such as GAM, in nuclear physics, particularly for complex, many-body problems where traditional methods face computational challenges.

Received Date
Accepted Date
Published Date

Contents

1 Introduction

Nuclear binding energy is a fundamental property that plays a crucial role in understanding nuclear stability, decay processes, and astrophysical phenomena. It reflects the energy required to disassemble a nucleus into its constituent protons and neutrons, providing insights into the forces that hold the nucleus together. Accurate prediction of nuclear binding energies is, therefore, essential for both theoretical research and practical applications in nuclear physics.

Traditionally, nuclear binding energies have been predicted using models such as the Liquid Drop Model (LDM). Although this model has been successful in many instances, it faces significant limitations, particularly when applied to nuclei that are far from the line of stability. The LDM, which treats the nucleus as a macroscopic liquid drop, approximates binding energy based on macroscopic properties such as volume, surface area, and Coulomb forces [1]. However, it struggles to accurately capture the finer details of nuclear structure, including the effects of magic numbers and other quantum phenomena. As a result, the Root Mean Square Error (RMSE) for LDM predictions typically reaches around 2.463 MeV, reflecting the model's limited accuracy [2].

Given these challenges, there has been growing interest in exploring alternative approaches that can offer greater accuracy and deeper insights into the underlying physics. Machine learning models, particularly those that can model non-linear relationships, such as the Generalized Additive Model (GAM), have emerged as promising tools in this regard. GAMs provide a flexible yet interpretable framework for predicting nuclear binding energies by capturing complex interactions between nuclear properties while allowing for the analysis of individual contributions of each feature [3].

In this study, we apply the Generalized Additive Model to predict nuclear binding energies, focusing on a set of nuclear features including proton number (Z), neutron number (N), and other key nuclear properties. Our goal is to not only achieve high predictive accuracy but also to provide interpretable insights into the role of each feature in determining the binding energy, thereby bridging the gap between traditional theoretical models and modern computational techniques. We compare the performance of the GAM against the traditional LDM to evaluate its effectiveness and to explore its potential advantages in nuclear physics research.

2 Methodology

2.1 Generalized Additive Model (GAM)

The Generalized Additive Model (GAM) is a flexible regression model that allows the response variable to be modeled as a sum of smooth functions of the predictor variables (features) [4]. Mathematically, the GAM can be expressed as:

$$g(E(Y)) = \beta_0 + f_1(Z) + f_2(N) + f_3(A) + \ldots + f_n(x_n) + \epsilon \tag{1}$$

where:

- $g(E(Y))$ is the link function that relates the expected value of the response variable Y (nuclear binding energy) to the linear predictor.
- $f_i(x_i)$ are smooth functions that model the effect of each predictor variable (feature) x_i on the response variable.
- β_0 is the intercept term.
- ϵ is the error term.

In this study, the identity link function was used, which means that $g(E(Y)) = E(Y)$, simplifying the model to:

$$E(Y) = \beta_0 + f_1(Z) + f_2(N) + f_3(A) + \ldots + f_n(x_n) + \epsilon \tag{2}$$

The smooth functions $f_i(x_i)$ were estimated using penalized splines, which balance the complexity of the model with the need to avoid overfitting. The model was implemented using the `pyGAM` library in Python [5], and the training process involved optimizing the smoothing parameters to balance bias and variance.

2.2 Data Source

The dataset used in this study is sourced from the Atomic Mass Evaluation (AME) 2020, which provides detailed information on nuclear properties [6]. The AME 2020 dataset includes various nuclear characteristics, such as proton number (Z), neutron number (N), mass number (A), and binding energy.

In this study, only data points where both neutron number (N) and proton number (Z) are greater than or equal to 8 were used. Data with $N < 8$ or $Z < 8$ were excluded from the analysis. This exclusion was made to ensure the model focuses on nuclei that are more relevant to the study of nuclear binding energy, as nuclei with very low N or Z do not provide significant insights for the predictive model.

2.3 Feature Engineering

The predictor variables (features) selected for this model include proton number (Z), neutron number (N), mass number (A), nuclear radius, surface area, $|N - Z|$, MagicZ, MagicN, $Z\_valence$, $N\_valence$, and pairing (Pair). These features were chosen based on their relevance to nuclear binding energy.

Proton number (Z) and neutron number (N) were directly obtained from the AME 2020 dataset as fundamental properties of nuclei. These are used to calculate the mass number (A), which is simply the sum of Z and N.

Nuclear radius and surface area were computed using empirical formulas related to the mass number A. Specifically, the nuclear radius was approximated using $A^{1/3}$, and the surface area was derived using $A^{2/3}$.

The feature $|N - Z|$ represents the absolute difference between the number of neutrons and protons in the nucleus, calculated directly from Z and N.

MagicZ and MagicN indicate the proximity of Z and N to the nearest magic numbers, which correspond to completed nuclear shells associated with greater nuclear stability. These magic numbers include 2, 8, 20, 28, 50, 82, and 126. The values for MagicZ and MagicN were calculated by finding the closest magic number to Z and N and recording the difference.

$Z\_valence$ and $N\_valence$ denote the number of protons and neutrons in the outermost shell, determined by comparing the proton and neutron numbers with the known magic numbers. For example, if Z lies between 28 and 50, the number of protons in the outermost shell is given by $Z - 28$.

Finally, the pairing (Pair) feature was determined using a custom scale that assigns a value from 0 to 9 based on the parity of Z and N. The specific values are detailed in Table 1.

2.4 Model Implementation

The Generalized Additive Model (GAM) was implemented using the `pyGAM` library in Python. The model was trained using the predictor variables (features) described above, with the goal

Table 1: Pairing feature values based on the parity of Z and N with respect to the relationship between Z and N.

	odd-Z odd-N	odd-Z even-N	even-Z odd-N	even-Z even-N
$Z > N$	0	3	5	7
$Z = N$	1	-	-	8
$Z < N$	2	4	6	9

of predicting nuclear binding energy. Penalized splines were used to estimate the smooth functions $f_i(x_i)$ for each feature, balancing model complexity and overfitting.

To ensure the robustness of the model, 5-fold cross-validation was employed during the training process. Specifically, the dataset was split into 5 subsets. The model was trained on 4 of these subsets and validated on the remaining subset. This process was repeated 5 times, with each subset serving as the validation set once. The average performance across all folds was used to assess the model's generalization capability and to fine-tune the smoothing parameters.

Finally, the model's performance was evaluated on a separate test set using the Root Mean Square Error (RMSE) to assess predictive accuracy. The interpretability of the model was also leveraged to analyze the contribution of each feature to the binding energy.

3 Results

The Generalized Additive Model (GAM) was used to predict nuclear binding energies using the selected nuclear features. The model achieved a Root Mean Square Error (RMSE) of 0.3 MeV, representing a significant improvement over the Liquid Drop Model (LDM), which had an RMSE of 2.463 MeV.

Figure 1 provides a comparison between the GAM and LDM. Panel (a) shows the distribution of binding energy differences between the LDM predictions and experimental results in the neutron (N) versus proton (Z) plane, highlighting substantial deviations. In contrast, panel (b) demonstrates that the GAM predictions are much closer to the experimental data across a wide range of N and Z, indicating better overall accuracy.

Figures 2 and 3 illustrate the neutron separation energies and two-neutron separation energies for isotopes of calcium (Ca), bromine (Br), tin (Sn), and lead (Pb). The predicted values align closely with the experimental data, demonstrating the model's capability to capture trends and subtle variations across different isotopic chains.

4 Discussion

The significant reduction in RMSE from 2.463 MeV with the LDM to 0.3 MeV with the GAM highlights the latter's ability to capture complex, non-linear interactions between nuclear properties. The comparison in Figure 1 indicates that the GAM provides a more accurate and reliable model, particularly for nuclei with magic numbers, where traditional models often struggle.

The neutron separation energies in Figure 2 showcase the GAM's effectiveness in predicting sharp discontinuities at magic numbers, which are indicative of closed nuclear shells. This ability to model shell effects accurately is further validated by the consistent predictions of two-neutron separation energies across various isotopes in Figure 3. The GAM's capacity to capture these subtle structural effects makes it a valuable tool for studying nuclear stability

and binding energy.

4.1 GAM Interpretations

The interpretation of selected nuclear features using the GAM model, as shown in Figure 4, provides valuable insights into the contributions of key variables to nuclear binding energy. For proton number (Z), the GAM model exhibits a quadratic pattern (Figure 4a), where binding energy increases with the number of protons until reaching a peak, then decreases. This pattern is due to the Coulomb energy within the nucleus, where an increasing number of protons enhances the repulsive force among them, thereby reducing the binding energy.

For neutron number (N), the GAM model (Figure 4b) indicates that binding energy continuously increases as more neutrons are added. Neutrons act as a balancing force to counteract the repulsive interaction between protons, thus enhancing the overall binding energy.

The positions of proton number (Z) and neutron number (N) at magic numbers (*magicZ* and *magicN*), and their valence (*Z_valence* and *N_valence*), show distinct patterns in the GAM model (Figures 4d-4g). Nuclei with atomic numbers in the third shell ($50 \leq Z \leq 81$) exhibit high binding energies, while those in the fourth shell ($82 \leq Z \leq 125$) display lower binding energies, as observed in elements like Sn and Xe [**?**]. This reduction is attributed to a lack of neutron pairing with protons, leading to increased proton-proton repulsion and, consequently, lower binding energy. For magicN, high binding energy occurs when $N \geq 126$, as additional neutrons provide an attractive force that balances proton repulsion, increasing stability.

For nuclear radius, surface area, and mass number (A), the GAM model shows a similar trend (Figure 4). Binding energy decreases to a minimum as the nuclear radius increases, then rises again. This pattern is mirrored in surface area and mass number, with minima occurring at radius ≈ 4, surface area ≈ 18, and mass number $A \approx 70$. The correlation among these features can be expressed as:

$$70 \approx 18 \times 4 \approx 4^3$$

This is due to the derivation of radius and surface area from mass number. Although these features are interrelated, they were included in the analysis to assess their individual effects, which the GAM model shows to be similar for all three.

The feature $|N - Z|$, representing the neutron-proton difference, suggests that nuclei with balanced neutron and proton numbers are more stable.

The pairing effect interpretation reveals that nuclei with odd-odd proton and neutron numbers have the weakest binding energy, while even-even configurations show the strongest binding energy. Odd-even configurations exhibit medium binding energy. This pattern highlights the stabilizing effect of nucleon pairing, where complete pairing in even-even configurations reduces the repulsive forces within the nucleus.

Further analysis and visualization of these features are necessary to gain a comprehensive understanding of their roles in determining nuclear binding energy. While the current study provides insightful interpretations based on the GAM model, it is not exhaustive. These findings contribute to the understanding of nuclear binding energy but indicate the need for more detailed research to fully explore the impact of these and other nuclear properties.

5 Conclusion

This study demonstrates the effectiveness of the Generalized Additive Model (GAM) in predicting nuclear binding energy, offering a significant improvement over traditional models such as the Liquid Drop Model (LDM). The GAM achieved a Root Mean Square Error (RMSE) of

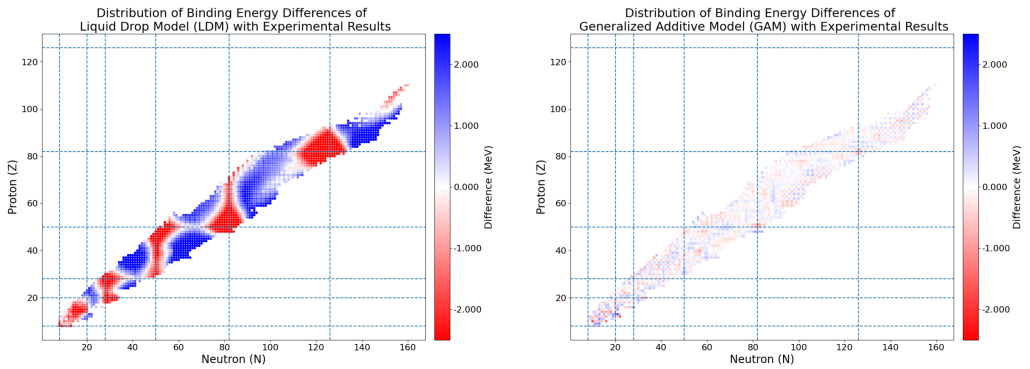


Figure 1: Comparison of binding energy difference distributions between experimental results and two models in the neutron (N) versus proton (Z) plane. The left panel shows the comparison with the Liquid Drop Model (LDM), and the right panel shows the comparison with the Generalized Additive Model (GAM).

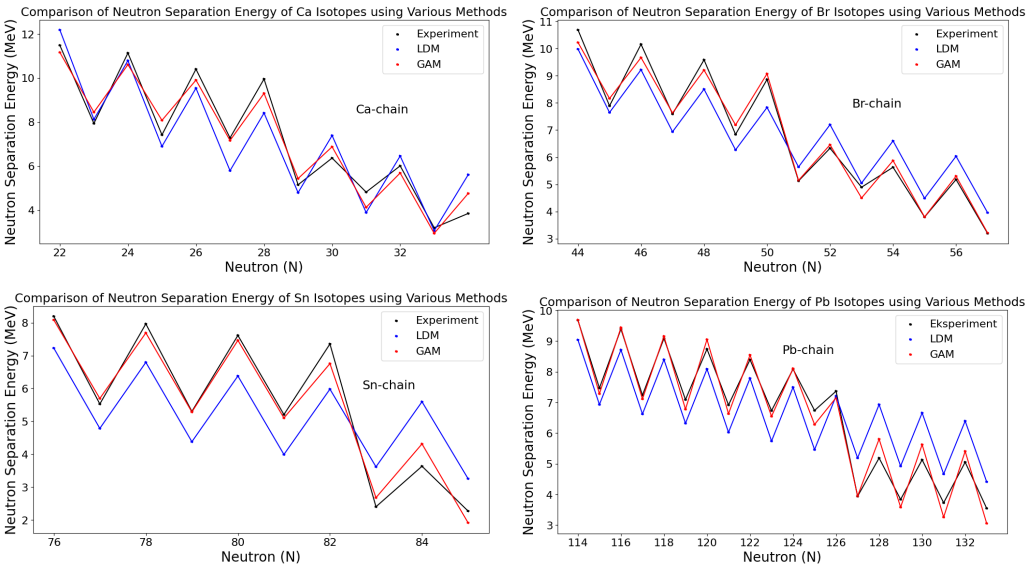


Figure 2: Neutron separation energy chains for the isotopes of Ca, Br, Sn, and Pb.

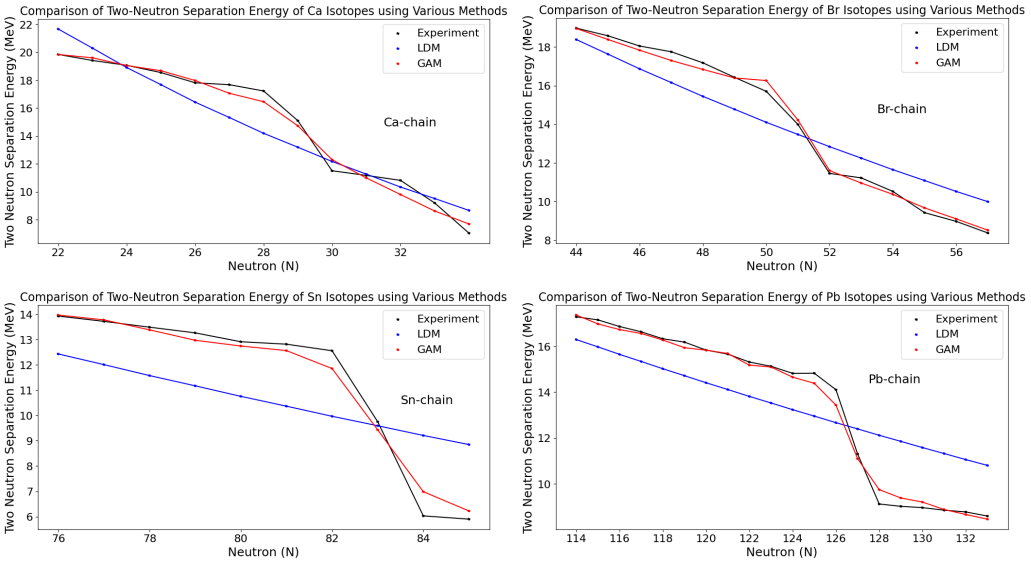


Figure 3: Two-neutron separation energy chains for the isotopes of Ca, Br, Sn, and Pb.

0.3 MeV, compared to 2.463 MeV for the LDM, highlighting its capability to capture complex, non-linear relationships between nuclear properties.

The model successfully identified critical patterns related to proton and neutron numbers, particularly at magic numbers, and accurately represented nuclear shell effects. Interpretations of nuclear features such as radius, surface area, and mass number revealed a coherent relationship, reinforcing the correlation between these variables and binding energy. Additionally, the analysis of pairing effects indicated that even-even configurations are the most stable, while odd-odd configurations exhibit the weakest binding energy.

These findings underscore the potential of machine learning models like GAM to enhance our understanding of nuclear binding energy and stability. However, further research and more comprehensive modeling are needed to explore the contributions of additional nuclear features and to refine the interpretations provided by this study.

Acknowledgements

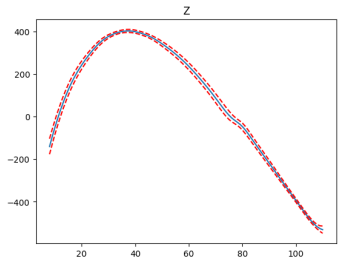
The author would like to thank Prof. Drs. Pekik Nurwantoro, M.S., Ph.D., for his guidance and support during this research.

A Supplemental Material

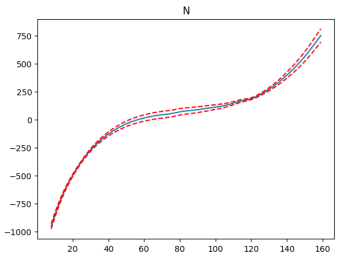
A.1 Data and Code

The dataset for the experimental binding energies used in this study is sourced from the Atomic Mass Evaluation (AME) 2020 and is referred to as `binding_energy_14.csv`. The dataset for the Liquid Drop Model is referred to as `LDM.csv`. The code used for analysis and visualization is implemented in Python. The data and code can be accessed at the following link: https://github.com/kristiyanlaoli/Predicting-Nuclear-Binding-Energy-Using-Generalized-Additive-Model.

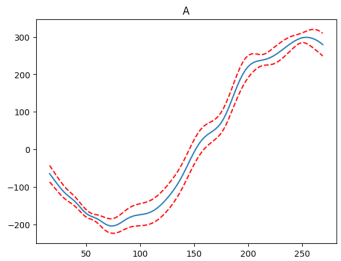
To replicate the results, users should ensure that all required libraries, such as `pyGAM`,



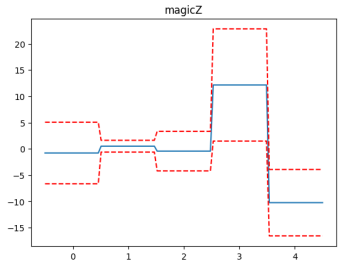
(a) GAM Interpretation of Z



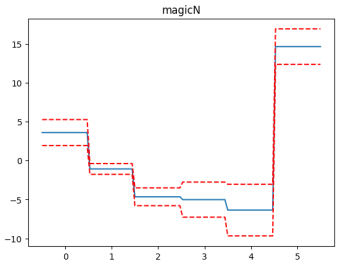
(b) GAM Interpretation of N



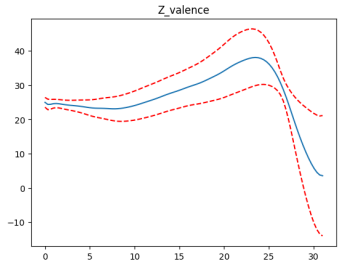
(c) GAM Interpretation of A



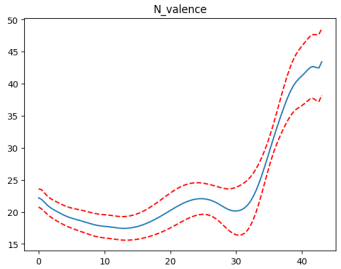
(d) GAM Interpretation of *magicZ*



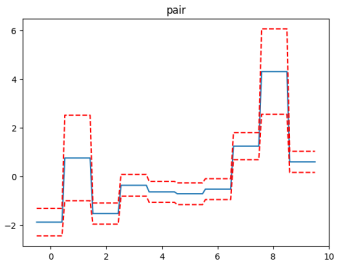
(e) GAM Interpretation of *magicN*



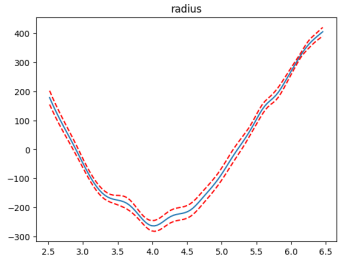
(f) GAM Interpretation of *Z_valence*



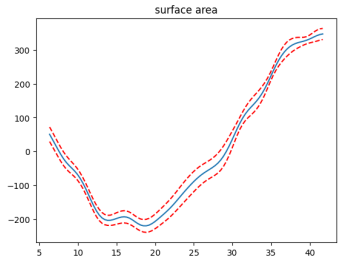
(g) GAM Interpretation of *N_valence*



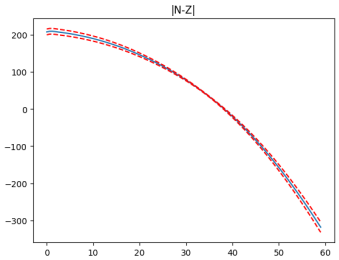
(h) GAM Interpretation of *pair*



(i) GAM Interpretation of *radius*



(j) GAM Interpretation of *surface*



(k) GAM Interpretation of $|N-Z|$

Figure 4: Interpretation of various nuclear features using the Generalized Additive Model (GAM).

`tensorflow`, `numpy`, `pandas`, and `matplotlib`, are installed. Instructions for running the code and loading the datasets are provided within the comments in the code.

References

[1] A. Bhagwat, *A simple model for nuclear structure*, Physical Review C **90**(6), 064306 (2014), doi:10.1103/PhysRevC.90.064306.

[2] Z. P. Gao, Y. J. Wang, H. L. Lü, Q. F. Li, C. W. Shen and L. Liu, *Machine learning methods in nuclear physics: A review*, Nuclear Science and Techniques **32**(1), 1 (2021), doi:10.1007/s41365-021-00956-1.

[3] J. H. Friedman, *Greedy function approximation: A gradient boosting machine*, Annals of Statistics **29**(5), 1189 (2001), doi:10.1214/aos/1013203451.

[4] T. J. Hastie, *Generalized Additive Models*, pp. 249–307, Routledge (2017).

[5] D. Servén and C. Brummitt, *pygam: Generalized additive models in python*, https://zenodo.org/record/2458509, doi:10.5281/zenodo.2458509 (2018).

[6] M. Wang, W. J. Huang, F. G. Kondev, G. Audi and S. Naimi, *The ame 2020 atomic mass evaluation (ii). tables, graphs, and references*, Chinese Physics C **45**(3), 030003 (2021), doi:10.1088/1674-1137/abddaf.